

(NASA-CR-142241) ON THE DEVELOPMENT OF AN
INTERACTIVE RESOURCE INFORMATION MANAGEMENT
SYSTEM FOR ANALYSIS AND DISPLAY OF
SPATIOTEMPORAL DATA (Texas A&M Univ.) 233 p
HC \$7.50 CSCL 05B G3/82

N75-19072

Unclas
12407

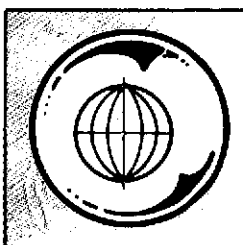
ON THE DEVELOPMENT OF AN INTERACTIVE RESOURCE INFORMATION MANAGEMENT SYSTEM FOR ANALYSIS AND DISPLAY OF SPATIOTEMPORAL DATA

By
John August Schell

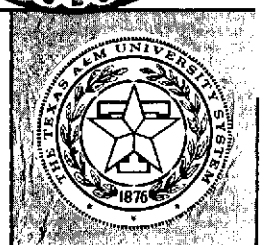
December 1974

Supported by
National Aeronautics and Space Administration
Grant NGL 44-001-001

RECEIVED
MAR 24 1 09 PM '75
OFFICE OF
UNIVERSITY AFFAIRS



TEXAS A&M UNIVERSITY
REMOTE SENSING CENTER
COLLEGE STATION, TEXAS



ON THE DEVELOPMENT OF AN INTERACTIVE RESOURCE INFORMATION
MANAGEMENT SYSTEM FOR ANALYSIS AND DISPLAY OF
SPATIOTEMPORAL DATA

By

John August Schell

December 1974

Supported by

National Aeronautics and Space Administration

Grant NGL 44-001-001

ABSTRACT

The recent availability of timely synoptic earth imagery from the Earth Resources Technology Satellites (ERTS) provides a wealth of information for the monitoring and management of vital natural resources. The quantity and rate of acquisition of this data requires that the resource planning function be aided by computer processing, however, computer software systems have not existed to provide the planning function with an adequate computer information structure to maintain image interpretations. Available systems also lacked the simplistic interactive interface required by the resource planner who is not versed in the complexities of computer processing.

Formal language definitions and syntax interpretation algorithms have been adapted to provide a flexible, computer information system for the maintenance of resource interpretation of imagery. These techniques are incorporated, together with image analysis functions, into an Interactive Resource Information Management and Analysis System, IRIMAS, which is implemented on a Texas Instruments 980A minicomputer system augmented with a dynamic color display for image presentation. A demonstration of system usage and recommendations for further system development are also included.

TABLE OF CONTENTS

<u>Chapter</u>	<u>Page</u>
ABSTRACT	ii
TABLE OF CONTENTS	iii
LIST OF FIGURES	viii
LIST OF TABLES	x
I. INTRODUCTION	1
An Information System	2
Objective	6
Scope of Report	7
II. BACKGROUND	9
Remote Sensing	9
Energy Mechanisms	12
Electromagnetic Radiation and Spectral Color	13
Sources of Energy	15
Atmospheric Attenuation	17
Signatures	22
Sensor Systems	26
Photography	28
Multispectral Scanner Systems	29
Microwave Sensors	32

<u>Chapter</u>	<u>Page</u>
ERTS.	33
Analysis Systems	34
Management Actions	36
III. INFORMATION SYSTEM DEFINITION AND SCOPE. . .	39
Introduction	39
System Design Criteria	41
Data Base Representation	43
Information Types	43
Raster Format and Display	45
Boundary Representation	47
Transformation and Data Handling	50
Coordinate Transformation	50
Data Transformation	51
Format Transformation	52
Analysis Operations.	53
Computer Aided Classification	54
Image Processing.	56
Communications	57
Output Products	57
Operator Communications	58
Error Diagnostics	59
IRIMAS--An Interactive Resource Information Management and Analysis System	61
IV. IRIMAS DATA HANDLING AND TRANSFORMATION COMPONENTS	63

<u>Chapter</u>	<u>Page</u>
Formal Languages	64
Language Terminology.	65
Syntax Driver	67
IRIMAS Information Syntax.	74
Block Label	75
Attribute Definition.	77
Data Entries.	77
Formal Syntax Definition.	80
Machine Format.	81
Data Syntax Conversion Subroutine . . .	83
Data Transformation.	93
Definition.	93
Conversion Subroutine	103
Transformation Interpreter.	103
Coordinate Transformation.	107
Transformation Definitions.	107
Syntax and Interpretative Subroutine. .	111
Data Display	119
Clipping.	121
Line Generation	125
Area Generation	127
Expansion Subroutine.	129
V. IMAGE PROCESSING	135
Image Analysis Functions	136

<u>Chapter</u>	<u>Page</u>
Smoothing	137
Edge and Line Enhancement	138
Density Slicing	139
Clustering.	141
Classification.	144
Image Processing Software.	146
VI. SYSTEM OPERATION	156
System Control	156
IRIMAS Input Drivers	159
INPUT	162
ERTS.	163
Processing	164
TRAIN	165
CLASSIFY.	166
CLUSTER	166
EDGE.	167
SMOOTH.	167
SLICE	168
TRANSFORM	168
XDATA	169
XCOORD.	169
EXPAND.	170
System Output.	170
DISPLAY	170

<u>Chapter</u>	<u>Page</u>
MAP	171
DEFINE.	171
HISTOGRAM, STATISTICS, DENSITY, OUTPUT.	172
Control.	172
VII. SYSTEM DEMONSTRATION	174
Information Processing	175
Operator Directives	176
Data File Definition.	178
Data Reconstruction	181
Brazos County Highway System.	184
Image Processing	191
IMAGE Operator Directives	196
Image Analysis.	196
Change Monitoring.	206
VIII. CONCLUSIONS AND RECOMMENDATIONS.	213
REFERENCES	220

LIST OF FIGURES

<u>Figure</u>		<u>Page</u>
II-1	Earth Observations Information System. . . .	11
II-2	Spectral Transmission of the Atmosphere. . .	20
II-3	Components of Radiance from a Natural Scene with 10% Reflectance	23
II-4	Reflectance Spectra for Natural Objects. . .	25
II-5	Multispectral Scanner Imaging.	31
III-1	Computer-processed Grey-scale Image of Brazos County.	49
IV-1	Logical Flow Diagram for Syntax Interpretation	71
IV-2	Logical Flow Diagram for SUBR.1.	90
IV-3	Logical Flow Diagram for SUBR.2.	101
IV-4	Coordinate Transformation Example.	108
IV-5	Logical Flow Diagram for SUBR.3.	114
IV-6	Logical Flow Diagram for TRANS	116
IV-7	Logical Codes for Clipping Algorithms. . . .	122
IV-8	Example of Line Rejection and Clipping . . .	124
IV-9	Example of Algorithmic Shading	128
IV-10	Logical Flow Diagram for EXPAND.	130
V-1	Logical Flow Diagram for PROCES.	149
VII-1	Example of Operator Dialogue	177
VII-2	Procedure for Definition of IRIMAS Data Files	179
VII-3	Traditional Insignia of Texas A&M University	182
VII-4	Operator Input Sequence for Computer Control	183

<u>Figure</u>		<u>Page</u>
VII-5	TAM Point Representation	185
VII-6	TAM Line Representation.	186
VII-7	TAM Area Representation.	187
VII-8	Operator Definition Sequence for Rotation. .	188
VII-9	TAM 45-degree Rotation	189
VII-10	General Highway Map of Brazos County	190
VII-11	IRIMAS Reconstruction of Highway System. . .	192
VII-12	ERTS Scene from Brazos County, Texas	194
VII-13	CRT Display System	195
VII-14	Operator Control Sequence for ERTS Processing	197
VII-15	Histograms of ERTS Image Bands	198
VII-16	Display of ERTS Image Bands.	201
VII-17	Operator Sequence for Classifier Training. .	203
VII-18	Classifier Training Results.	204
VII-19	Operator Specification of Classify Sequence.	205
VII-20	Classification Results for "Water" Class . .	207
VII-21	IRIMAS File Listing of Brazos River Data . .	208
VII-22	ERTS Image Sequence During Highway 6 West Bypass Construction.	209
VII-23	New Highway Construction Update to General Highway Map.	212

LIST OF TABLES

<u>Table</u>		<u>Page</u>
IV-1	Example Syntax Table	68
IV-2	Information Syntax	85
IV-3	SUBR.1 Operation Codes	88
IV-4	Data Transformation Syntax	97
IV-5	Syntax Conversion Subroutine SUBR.2 Operation Codes.	104
IV-6	Subroutine SUBR.2 Interpretation Functions .	105
IV-7	Polish String Operation Codes.	106
IV-8	Syntax for Coordinate Transformation	112
IV-9	Syntax Conversion Subroutine SUBR.3 Operations	120
V-1	Operation Codes for Enhancement and Classification	148
VI-1	IRIMAS Operator Directives	161

CHAPTER I

INTRODUCTION

Humanity is consuming ever greater quantities of land and natural resources; a situation which is causing great concern among many groups within our society. In this "land of plenty", shortages in desired commodities have already become apparent and the efficiency of the conversion of resources to the "good life", and even the "good life" itself, are being questioned and criticized. Modifications in the nominal mode of living in the United States, as in many parts of the world, are being dictated by resource shortages, public opinion, and legislative reaction. A major emphasis is being placed on the development and utilization of resource management systems to provide for more efficient conversion of valuable resources, improved agricultural production, and minimization of the effects of blight on the environmental and social systems.

A key element in the development of resource management systems is the acquisition and utilization of large quantities of information which is varied in nature and areal extent. Consequently, Earth Observations has emerged as an important field in the support of resource management. The techniques of acquiring the regional information have

This dissertation follows the style of the Proceedings of the IEEE.

developed to the point that resource managers are now being overwhelmed by the great quantities of raw data generated by current technology. As a result, it is important that the technical knowledge developed by practitioners of the art and science of remote sensing be distilled into an analysis/display system amenable to utilization by resource management agencies at local and regional levels. It is only in this way that the vital information contained in a synoptic and timely overview of our planet can be applied immediately and specifically to the problems facing our society as a whole.

It is the objective of this research to develop an information management and analysis system suited to the general needs of resource managers. The information system described implements the analysis techniques of the Earth Observations community in a minicomputer software/hardware system developed to provide a convenient, interactive interface between earth observations data and its users, and to implement a flexible multipurpose data handling scheme suited to handle the spatial and temporal information required for resource-related decision making.

An Information System

Several farsighted authors [1], and [2] have described the salient features of a computer land-use information

system. Such an information system could as easily apply to a broad range of resource management information requirements. The foremost requirement of such an information system must be the effective and efficient presentation of data from multiple sources in the form of products which are relevant to the problems and processes of resource management.

According to Brooner and Nichols [2],

If any method is to be successful, it must meet the criteria of timeliness, accuracy, and economic feasibility. Data must be recorded, massaged, and retrieved in a manner that local and regional agencies that most desperately require such information can effectively utilize it.

Condensing the thoughts of these authors, resource management information systems should consist of the following components: (1) system management, (2) data files, (3) interpretation/analysis, and (4) output product.

The system management component provides a house-keeping/input function. This function provides for the input of data and its transformation into formats suitable for insertion into data files. It translates data from remote sensing imagery, surveys, and administrative records into machine records. Additionally, the system management component includes the software monitor of the total system, and constitutes the input interface to the resource management activity.

The data file component separates data into disaggregate data, areal-unit summaries, and time-period summaries. The disaggregate data, which includes satellite and aerial imagery, survey information, etc., is temporarily stored information related to a common geographic base. Areal-unit summaries, and time-period summaries include geographically referenced resource information and other generally related parameter information which is further analyzed and displayed in a spatial format suitable for individual interpretation. These information files also provide the necessary data for the compilation of area statistical summaries.

The interpretation/analysis component functions are not fully defined by Dueker and Horton [1], and Brooner and Nichols [2], but both sets of authors have recognized the ultimate necessity of these general functions in a resource management system. Swanlund [3], and Viglione [4], have described some of the current techniques utilized in interpreting and handling image and spatial data. Numerous other authors [5], [6], [7], have described the theory and techniques of image enhancement and thematic classification from image survey systems.

Additional authors have described statistical model analysis procedures for the interpretation and prediction of urban change and urban transportation requirements, but

many of these studies are limited in scope and interpretability because of the immense information management problem. Harris [11] and Candeub [12], have indicated that the development of computer information management systems and more adequate display of spatial information hold the key to further development in urban modeling and its use in the planning function. Many systems engineering study groups have recently become active in the application of systems analysis techniques to social and resource problems [13], [14]. Yet again, the key to meaningful results is the management of the spatial information required by the models.

One of the most significant aspects of an information management system is its output product, for it is upon this component that the whole system must be judged. With these words,

In terms of land resource management, the time-space aspect of the output product cannot be overemphasized. Ecology also requires graphic and cartographic methods especially suited to analysis of spatial data.

Brooner and Nichols [15] emphasize the output requirements of thematic maps and statistical displays, and Thomlinson [16] says that these type products are indispensable for ecological analysis.

The works of many authors have served to provide a strong conceptual definition of a computer resource

information management system, and have provided the inputs from practitioners of the science of urban land management required for the engineering of such a system. Many other authors have provided the techniques required for the automatic and manual analysis of spatial and temporal inputs to a resource information management system afforded by the science of earth observations and remote sensing. Recent hardware developments in minicomputer technology and low-cost display systems have essentially provided the raw materials required for the development of a computer information management and analysis system amenable for utilization by a resource management function.

Objective

It is the objective of this research to design and develop an Interactive Resource Information Management and Analysis System (IRIMAS), configured for simplistic user interaction and cost-effective implementation. This system will facilitate integration of remote sensor surveys and will provide access to this composite data base through spatial display media. Techniques for information enhancement and class identification will be implemented and included in the system to provide a versatile management tool.

Scope of Report

The remote sensing field is varied and covers broad interdisciplinary areas. The specific equipment, techniques and physical considerations with regard to energy mechanisms which are being measured; the analytical and automated methods of analysis of data; and the application of results to diverse activities all tend to discourage a cohesive understanding of the techniques and philosophies involved. Chapter II is written to give a general overview of the field of remote sensing and remote sensing activities covering energy mechanisms and spectral distributions, sensor systems in operational use, and philosophy of remote sensing analyses. It also suggests the needs of resource managers in the application of remote sensing techniques to problem solving.

A resource information management and analysis system is described in a general manner in Chapter III. Format difficulties with current analysis systems and desirable characteristics of a comprehensive information system are described, and a modified format for resource information entry and retrieval are defined. Chapters IV, V, and VI discuss in technical detail the Interactive Resource Information Management System (IRIMAS). A demonstration of the Interactive Resource Information Management System utilizing data available from Brazos County, Texas, is

discussed in Chapter VII, while conclusions and recommendations for system improvements are covered in Chapter VIII.

CHAPTER II

BACKGROUND

Remote Sensing

Remote Sensing is an aggregate of activities concerned with inferring the nature of objects or areas by their measurement at a distance. These activities embody the physical sciences concerned with energy mechanisms, since it is generally changes in energy which are measured; the engineering sciences concerned with the detection and recording of the energy changes; and the analytical sciences, including the activities of mathematics, statistics, information and computing sciences; photogrammetry and image interpretation; and the other sciences which transform the recorded measurements. Also included are the application disciplines such as agriculture, geology and sociology which combine discipline experience with the information provided by the other groups to make the required inferences. The remote sensing field has evolved from a forced intermingling of these diverse activity groups; each with its own terminology, techniques, philosophies, and performance evaluation criteria. The worth of the composite remote sensing activity is judged by the social and economic impact of its products.

Remote sensing activities have been modeled by Rouse

and Schell [17] in terms of a large scale information system directed toward the provision of specific application products. This system can be viewed in terms of the four functional areas represented: Energy Interactions, Energy Sensing Systems, Analysis Activities, and Management Evaluation and Action (See Figure II-1).

The remote measurement of environmental characteristics is directly related to the measurement of energy surrogates of those characteristics. Energy released by a variety of mechanisms is interacted with by the Environmental Situation in a characteristic fashion. To the extent that these energy interactions are unique for a particular environmental situation, a knowledge of the character of the energy before the interaction and a measurement of its characteristics after the interaction can lead to a knowledge and characterization of the situation which produced the changes.

Well developed remote sensor systems are available for the routine measurement of energy. Some of these sensor systems have been applied to a general data gathering mission and have provided a variety of measurements of the environment in a timely and systematic manner.

Similarly, data handling and analysis techniques have been developed along well specified patterns and form identifiable groupings. It is primarily in the provision

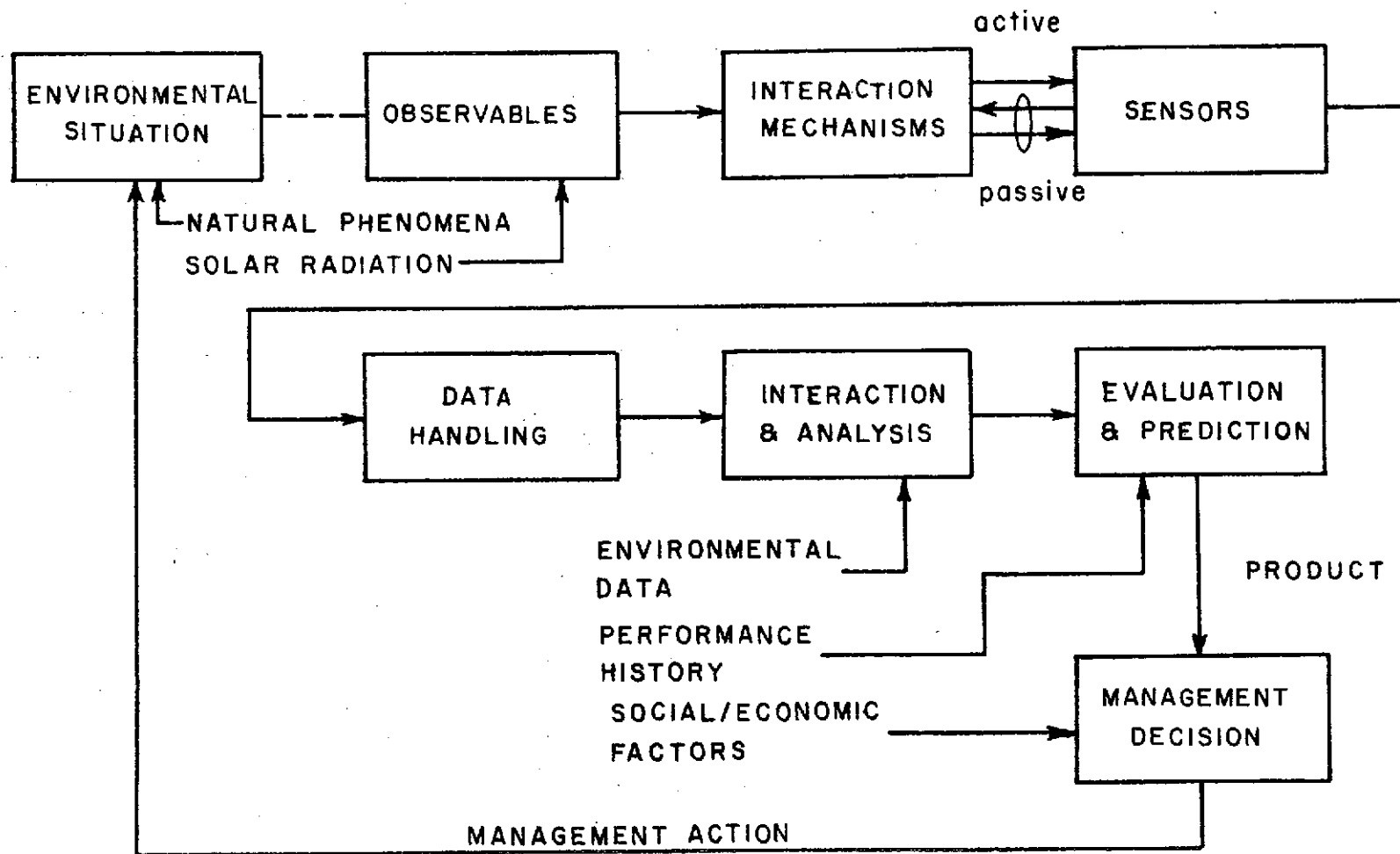


Figure II-1. Earth Observations Information System

of data products to the Management Evaluation and Action activity and the specific products useful to an environmental manager that the system is ambiguous. As such, the output data products produced by the analysis activities tend to be self-serving rather than fully beneficial to the next function block in the earth observations system.

Energy Mechanisms

The primary source of information available in remote sensor measurements is the spectral distribution of reflected or emitted energy. Some comprehension of the sources and change characteristics of energy is necessary to obtain a fuller understanding of the measurements of remote sensing and their applicability in specific problem solving exercises. The full spectral range of energy from light well into the microwave region is measured by remote sensors, however, the majority of resource characterization by remote sensing techniques has been accomplished with the help of visible and near-visible radiation. It is this spectral region which is receiving primary emphasis, although microwave measurements and interpretation offer great future promise.

Fundamental to an appreciation of energy changing mechanisms in the visible and near-visible regions is a knowledge of solar radiation and the atmospheric effects of

absorption and scattering. The characteristics of vegetation, soils, and other natural resources which change the energy spectral distribution are important to the successful application of automatic classification algorithms, and also to manual interpretation of sensor measurements. The total energy change is characterized by a spectral signature which is the measured energy distribution influenced by the sensor target. Within the uniqueness of the characteristics of this signature often lies the success or failure of the measurement interpretation task.

Electromagnetic Radiation and Spectral Color

Two common models are utilized to describe electromagnetic radiation, the wave model and the particle model. According to electromagnetic wave theory, energy is propagated through space in a wave pattern analogous to sound waves in the air or shock waves in water. An electromagnetic wave consists of two fluctuating fields which are generated as a result of an electrical charge being accelerated and the wave frequency depends on the number of accelerations per second. The second approach to the study of radiative energy transfer is explained by particle or photon theory. This theory states, in effect, that energy is transferred in discrete units called quanta or photons and that the energy of a photon varies directly

with frequency. Both theories are consistent and serve to provide a complementary understanding of electromagnetic energy radiation over the widest spectral range.

The term "spectrum" has been applied to the continuous series of colors obtained when white light is separated into its component colors of red, orange, yellow, green, blue, indigo and violet. This series taken together is called the visible spectrum. On the basis of the wave theory of light, color is associated with the radiation wavelength. In the visible spectrum, red has the longest wavelength and violet has the shortest wavelength. The visible spectrum is only a small portion of a much broader band of radiation exhibiting properties similar to those of visible light. This total range of radiation is called the electromagnetic spectrum. This enlarged spectrum definition includes radiowaves, microwaves, infrared or heat rays, visible light, ultraviolet rays, X-rays, and Gamma rays, although in remote sensing the utilization of microwave, infrared, visible and ultraviolet radiation is more significant.

Variations in color represent the interpretation that man's eyes and brain make of the differences in wavelength or frequency of light. Color is always associated with the way things are seen. Two simple concepts help define the color of an object. If various colors of light are

allowed to fall on a white screen, the result is that each color is faithfully reflected, indicating that a white object reflects all colors or wavelengths. If white light is allowed to fall on a colored surface, say red, the surface absorbs all the colors except red, which is reflected. Therefore, the color of a material can be defined as that of the light which it reflects.

Absorptance is that quality or characteristic of an object which allows it to capture photons of selected wavelengths, while the reflectance of an object causes certain wavelengths to be re-radiated. A third characteristic affecting radiation is transmittance which is that property or characteristic of an object which allows the photons to pass unaltered through the object.

Sources of Energy

Sources of energy radiation may be generally categorized into two types, broadband radiators and those producing line spectra or monochromatic radiation. For the most part, sensing in the visible and near-visible portions of the spectrum relies on broadband illumination or radiation. Solar radiation, a primary visible wavelength source, is broadband. It is only at the level of an individual molecule that line spectra radiation and absorption become important. Laser illumination is of the single

wavelength variety and sensing a laser illuminated scene also requires special consideration.

All objects at temperature above absolute zero (0°K) radiate energy. The theoretical limit of the spectral distribution of this radiation is described by Planck's law;

$$\mu_{\nu} = \frac{8\pi h\nu^2}{c^3(e^{h\nu/kT}-1)}$$

where μ_{ν} is the amount of radiant energy per unit frequency interval and ν and T are the frequency and temperature, respectively. The constants c , k , and h are the speed of light, the Boltzman constant, and Planck's constant. This radiation is characterized by an object which absorbs radiation equally as well as it radiates and with the same spectral weighting or distribution. Emission at the theoretical limit where a substance is a perfect absorber and emitter of energy is referred to as blackbody radiation.

No real body is a perfect emitter, although some substances, such as lampblack or powdered graphite come close. They fail because they do not absorb or emit equally over a wide spectral range. The ratio of the emittance of a real object to that of a blackbody is called its emissivity and is a measure of the radiating efficiency of the object. The emission of real bodies behaves similarly to the blackbody laws in that the hotter the body, the more

the radiation and the shorter the dominant wavelength. However, the differences in the characteristic emission of a blackbody and that of a real body may be markedly different in certain spectral bands.

Atmospheric Attenuation

Since the nature of remote sensing is the measurement of objects at a distance, the effects of the intervening distance on the transmitted energy is important. In addition to the reduction of power density as an inverse function of the distance between the source and the detector, the phenomena associated with atmospheric attenuation must be considered. Two major mechanisms are interacting with the energy flux density resulting in its attenuation. Specifically, these mechanisms are atmospheric absorption and atmospheric scattering.

Beer's Law is a relationship which describes the change in flux density of a plane-parallel beam of monochromatic radiation as it penetrates a medium. This law is written:

$$I_{\lambda}' = I_{\lambda} e^{-\alpha_{\lambda} x}$$

where I_{λ}' is the resultant flux density from the initial flux density I_{λ} attenuated by the exponential factor with parameters α_{λ} , the attenuation constant in nepers/meter, and x , the distance in meters through which the wave has

traveled. The distance x through which the wave must travel to be attenuated 37.3% or $1/e$ is referred to as the optical depth and has a value of $1/\alpha_\lambda$.

A major contribution to atmospheric attenuation, particularly in the visible and near-visible region of the electromagnetic spectrum, is molecular absorption by atmospheric constituents. It is important to realize that attenuation is related to both the rate of attenuation α_λ and to the distance through which the energy must travel, x . In certain circumstances where the distance between sensor and target is small, the effects of the atmosphere will also be small. However, for satellite measurements where atmospheric absorption is high, the effects of the atmosphere can no longer be summarily neglected.

In the ultraviolet portion of the spectrum, the nitrogen and oxygen absorption of radiation in the region from $0.13\text{-}0.22\mu$ is nearly complete. Radiation at $0.22\text{-}0.30\mu$ through an ozone layer is also completely blocked. Oxygen absorption bands occur at 1.06 and 1.27μ in the infrared portion of the spectrum, while many absorption bands of water and carbon dioxide along with nitrous oxide, methane and ozone occur in the $1\text{-}24\mu$ region.

The atmosphere is completely opaque in many portions of the infrared spectrum. Absorption by atmospheric water vapor virtually closes the atmosphere from $25\text{-}1000\mu$. Other

water vapor absorption characteristics of the atmosphere is found at 2.5mm, 5.0mm and 1.35cm. Despite these absorption characteristics of the atmosphere, many "windows" remain where the atmosphere is nearly transparent, and a significantly large percentage of energy is allowed to pass. Thus, the observation of the interaction of impending energy with a target is limited and in some sense is determined by the transmission characteristics of the atmosphere in the spectral band of interest. In Figure II-2 is shown a graph of atmospheric transmission in various spectral bands.

Aside from atmospheric absorption, there is also the problem of scattering by particles, water droplets and clouds within the atmosphere. Scattering is the process by which small particles suspended in a medium of different index of refraction diffuse a portion of the incident radiation. In most scattering (Raman being the major exception) no energy transformation results, only a change in the spatial distribution of the radiation results. Scattering is often a major cause of attenuation in the atmosphere.

Scattering is a function of (1) the ratio of the particle diameter to the wavelength of the radiation; (2) the nature of the composition of the particles as characterized by their complex index of refraction;

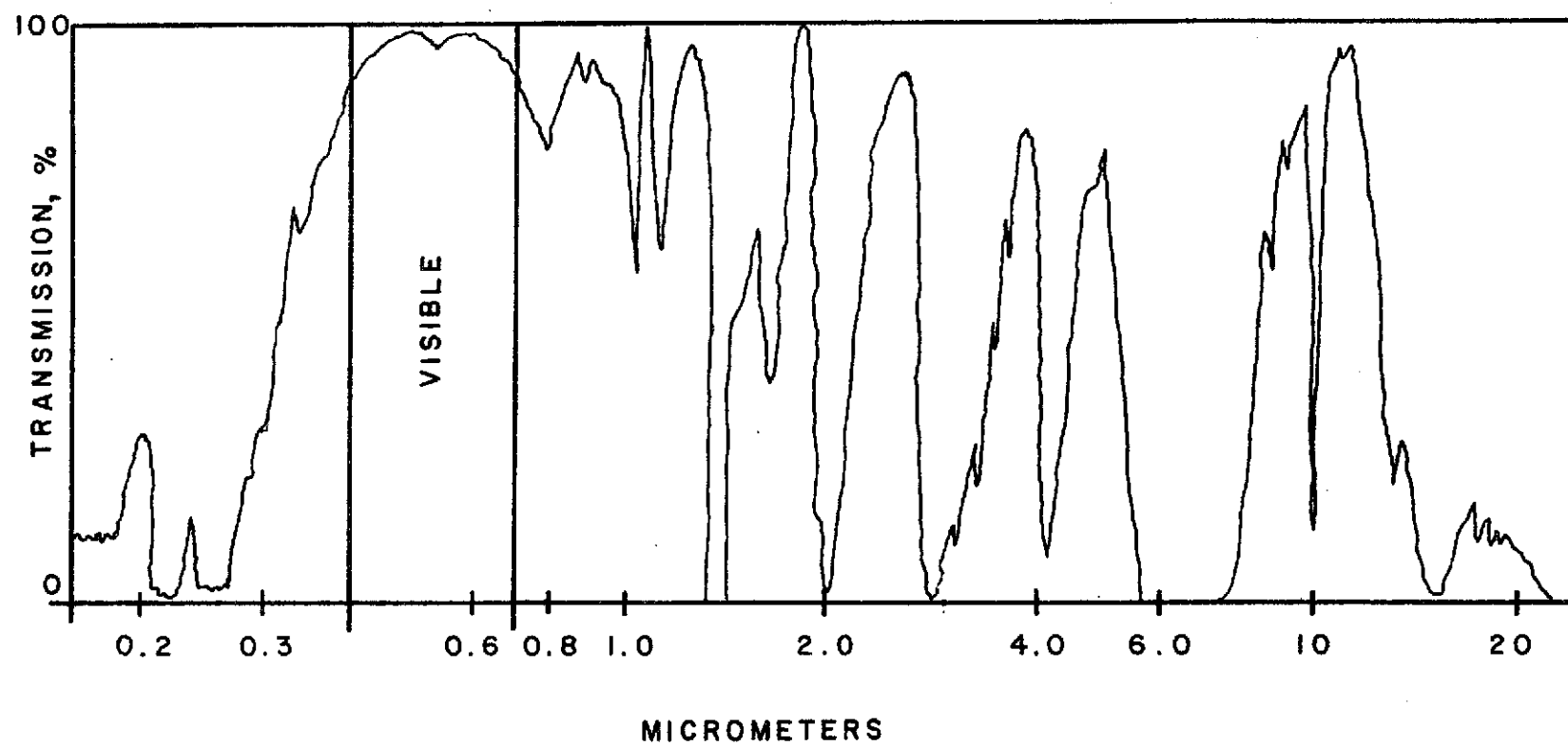


Figure II-2. Spectral Transmission of the Atmosphere

(3) the particle size and number distribution; and (4) the shape of the scattering particles.

When the size of the particle diameter is on the order of magnitude of about a tenth of a wavelength, scattering occurs in general accordance with the Rayleigh scattering law. According to the Rayleigh scattering law, scattering, hence attenuation, varies approximately with the fourth power of the wavelength. For larger particle sizes, on the order of magnitude of a wavelength, the scattering varies in a complex fashion as described by Mie theory. Mie scattering is characterized by re-radiation in the direction of the primary wave. Since the re-radiation of secondary waves is superimposed on the incident plane waves, not all of the energy is lost from transmission in the direction of the primary wave. A general relationship governing Mie scattering which is applicable to atmospheric aerosol in the visible and near infrared regions takes the form of $S = C_2 \lambda^n$, where the exponent n , which is a function of particle size, varies between -4 and a small positive number. At particle sizes ten times the wavelength, non-selective or wavelength independent scattering occurs. This particle size scattering effect generally serves to mark the somewhat diffuse upper bound of the realm of scattering and gives way to the laws of geometrical optics.

Another form of scattering which is present, but

relatively small in comparison, is "Raman" scattering. In this type of scatter there is an inelastic collision between a photon and a molecule. The result is a change in the energy of the photons and consequently its wavelength. Thus the photon, in addition to being scattered, has a modified wavelength which is dependent upon the energy state of the molecule. The importance of the "Raman" affect is that it may be used to distinguish composition and concentration of the molecular makeup of the atmosphere under special circumstances, but has a relatively miniscule affect when contrasted to other scatter mechanisms.

Signatures

Three basic energy interaction mechanisms are monitored by remote sensing instruments. These mechanisms include energy emission, absorption and reflection. Natural radiation is governed by Planck's law which describes the ideal emission from objects at temperatures above absolute zero (0°K). In Figure II-3 are shown the emission spectra for solar radiation at 3000°K and for emission from objects at 300°K , the average temperature of the earth. These curves show the dominance of solar radiation in the visible portion of the spectrum ($0.4\text{-}0.75\mu$) and the near infrared portion ($0.75\text{-}3.0\mu$).

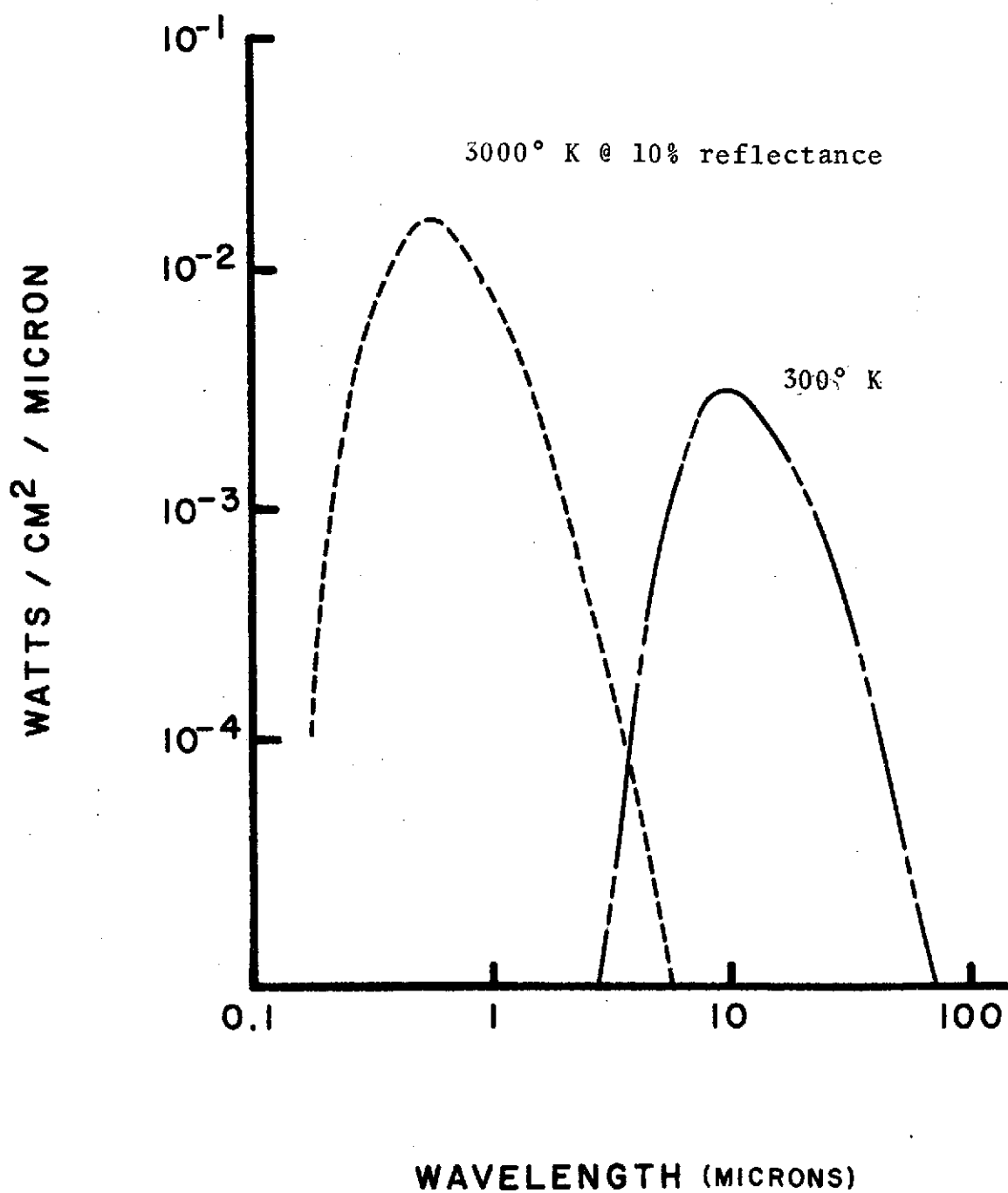


Figure II-3. Components of Radiance from a Natural Scene with 10% Reflectance

The radiation from the earth at 300°K becomes dominant in the thermal infrared portion of the spectrum (3.0-14.0 μ).

Solar radiation in the 0.4-3.0 μ portion of the spectrum masks the radiation contribution of the earth even for low surface reflectances, therefore, it is this surface characteristic which most influences the energy measured. In this region, the reflectance and absorptance are for the most part complementary characteristics, and those substances or mixtures which have unique reflectance and absorptance spectra can be identified. The term "signature" has been commonly applied to spectral reflectance measurements of objects and characterizes a specific reflectance-absorptance relationship. In Figure II-4 are shown typical reflectance spectra for some natural objects.

The radiation from earth targets at 300°K becomes dominant in the region of the spectrum from 3.0-14.0 μ . It is in this region that the target temperature and emission characteristics strongly influence the measurements made. Measurements from the thermal infrared portion of the spectrum provide information about the temperature of an object.

Beyond the thermal into the longer wavelengths lies the microwave region of the spectra. This region is more strongly influenced by the electromagnetic properties of volumes rather than strictly surface characteristics or

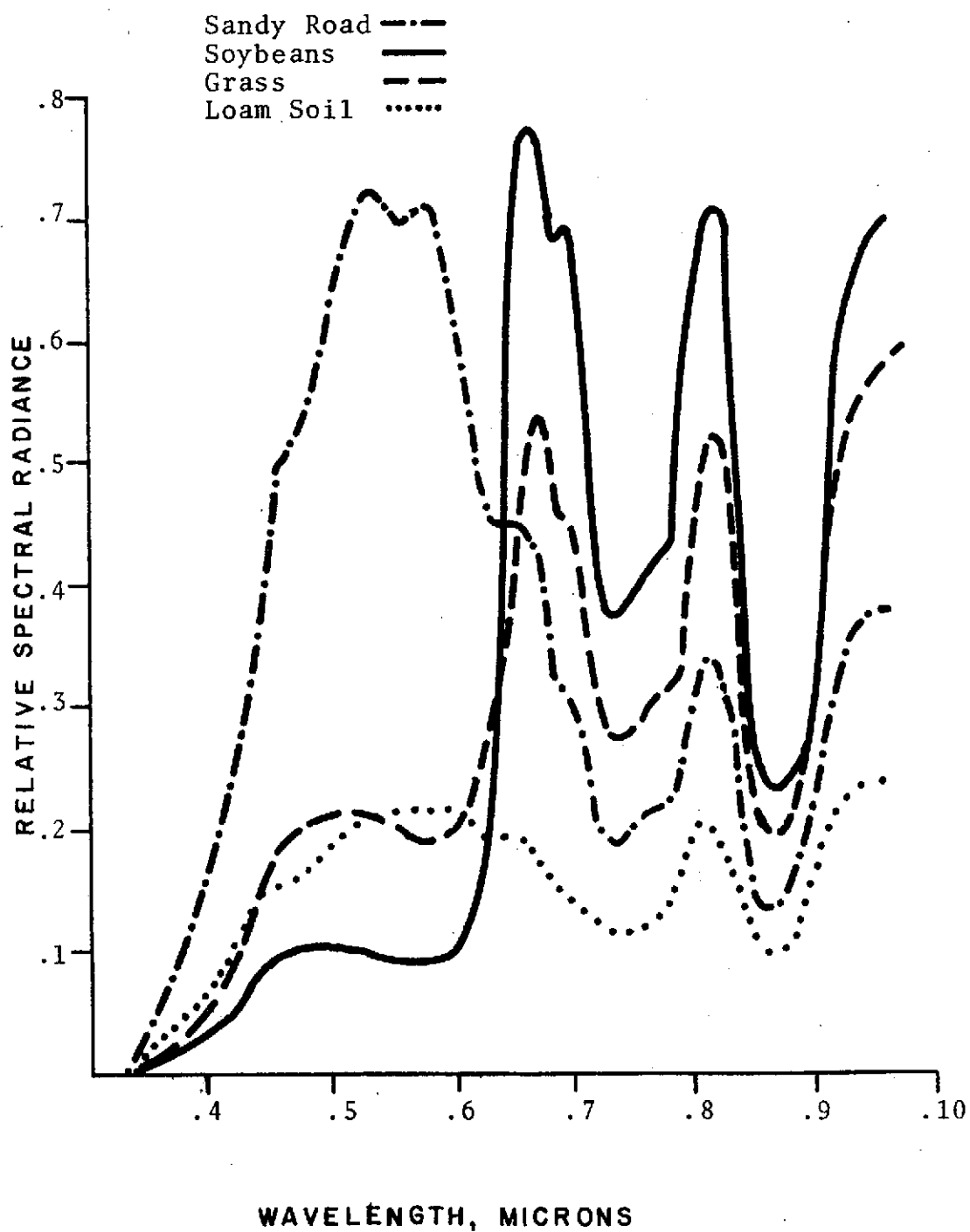


Figure II-4. Reflectance Spectra for Natural Objects

thermal perturbations, and may be sensed through passive radiometer measurements or by active radar techniques. Target signatures also exist in this spectral region which relate more strongly to the actual material composition of objects rather than to their surface characteristics alone. Although microwave sensing is rapidly developing, the remote sensing field has been largely dominated by visible and infrared sensing and these portions of the spectra have provided the bulk of the available measurements.

Sensor Systems

Sensor systems used in earth resources survey are varied in their techniques of energy measurement, the portion of the spectrum in which the energy is measured, and the ultimate format of the measurement presentation. Sensors may be described by their respective spectra, visible, infrared (reflective and thermal), and microwave; or they may be referenced by their format of presentation, either image or linear. They may be classed as either active or passive depending upon whether energy is radiated by the sensor and the return measured or whether the energy sources are external to the sensor and only their effects are measured.

The most common sensor system is the camera. This imaging sensor is passive and photographically records

reflected energy from surfaces illuminated by an external source, typically the sun. Photographic sensing measures energy in the visible and reflective infrared portions of the spectrum ($0.4\text{-}1.0\mu$). This sensing is either broadband, where the energy density is integrated spectrally over the spectral range of the system (polychromatic black and white), or the energy in spectral bands may be measured by selecting appropriate band pass optical filters and film combinations. Color film duplicates the spectral filter/film combination wholly within the film.

Other visible imaging systems belong to the multispectral scanner class of instruments. Images are formed by multispectral scanner (MSS) systems by making measurements from an optical system which scans the area to be imaged. This instrument is passive and may be used to record the energy in several spectral bands simultaneously from a single target area. The multispectral scanner technique is also used to extend imaging into the reflective and thermal infrared portions of the spectrum by an appropriate choice of system detectors.

Sensor systems also exist which provide image sensing in the microwave region. These include both the active radar sensors and the passive microwave radiometer systems which provide a unique view of the earth.

Photography

Photography is an important part of remote sensing activities. Not only are direct photographic recordings of remotely located objects frequently employed, but photography may also be the medium in which other sensor information is displayed. The principal advantage of the photographic process is its ability to quickly and accurately record radiant energy, and to do so economically. In order to utilize the photographic system to its full value, an understanding of the properties and limitations of the photographic process is necessary.

Photographic film consists of a thin layer of base material, one or more emulsions and a subcoat which causes the emulsion to adhere to the base. The photographic emulsion is a specially prepared gelatin containing the photosensitive materials in suspension. Most emulsions also contain dyes to affect the photochemical process. Color film utilizes the basic photographic emulsion in three layers to produce the color image. Since the principal intention of color film is to replicate what the eye sees, only three dyes and three emulsions are necessary. An additional dye and emulsion layer would not be useful for normal viewing.

In ordinary color film, the reception of blue light activates the appropriate dye system to replicate the blue

appearance by the processed film. Similarly, the received green light activates the dye system for replicating green and likewise for the red. In a color translation system such as infrared film, the received color activates a different dye system. In the case of infrared film, the blue light is filtered, the green light is made to produce blue on the film, yellow and red produce green and the received near-visible infrared produces the red excitation from the film. Thus, infrared reflection from an object is recorded and presented as red in the developed infrared film.

Multispectral Scanner Systems

Electromechanical scanner systems are imaging devices in which the scene to be mapped is physically scanned by a telescope system and the energy collected is converted to an electrical signal by a detector or detector array [18]. This method of image production has provided the majority of multispectral imaging available to date. The technique is reliable, flexible and proven. A considerable amount of development is being undertaken to replace the electromechanical systems with other electronically scanned devices such as the return beam vidicon, photodetector arrays, and the image dissector tube, however, these systems have not been adequately developed or proven to be fully

operational. The eventual replacement of the electro-mechanical scanner systems will not, however, change the basic sensing characteristics or image presentation modes produced by this general class of instruments.

Scanner systems produce images by sweeping the detector image from side to side across the scene to be imaged while the scanner vehicle moves forward. This raster method of image scanning is graphically displayed in Figure II-5. The multispectral scanner technique also filters the monitored radiation into spectral bands prior to detection. Detection of the radiant energy in each of the spectral bands produces multiple images of the scene, which can be analyzed according to the individual spectral distribution of individual picture elements. The scanner system is an improvement over standard aerial photography in that both the spectral resolution and the absolute radiance calibration are more exact. With the magnetic tape recording of the signals produced by the scanner detection system, the data are in a format that is more amenable to automatic processing either by analog or by digital techniques. An additional advantage of the scanners is the fact that detectors are available for imaging outside the normal visible and near infrared ranges of photography.

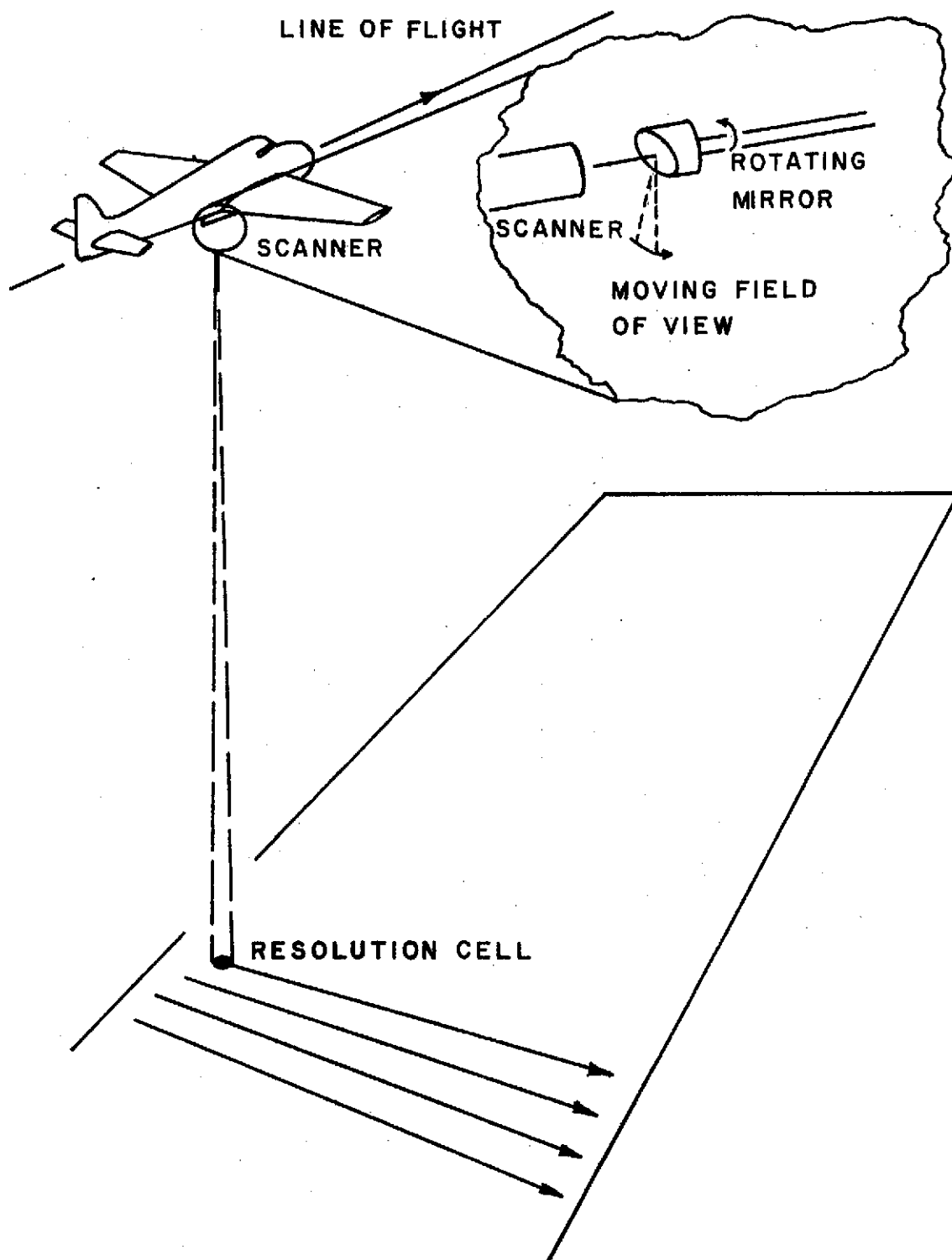


Figure II-5. Multispectral Scanner Imaging

Microwave Sensors

Other imaging systems widely in use are the side-looking radar systems. These radar systems produce images by transmitting microwave pulses transverse to the direction of flight. The radar return amplitude is measured as a function of time and is traced across a photographic film by an amplitude modulated glow lamp or similar device. This technique produces an image similar to the scanner system, however, the energy pulse transmitted is controlled so that an absolute measurement of the energy return is possible. The scanner system relies generally on the unknown radiative characteristics of the sun or of the terrain imaged. Synthetic aperture antenna techniques and other enhancement procedures may be utilized to obtain fine resolution and high image quality from the side-looking radar system.

Electronically scanned array antennas may also be utilized with sensitive, low noise, broadband microwave receivers to produce microwave images. These images are produced as a result of measurement of the emissive characteristics of the terrain in selected microwave bands. Each of these microwave sensors provides additional knowledge of the terrain in a portion of the spectrum which has been relatively ignored. As the microwave characteristics of natural terrain features continue to be

investigated, great promise is shown for obtaining data with as high an information content as is currently obtained in the visible and near-visible region. Promise is also high for obtaining a better view of subsurface phenomena which cannot be directly detected by other than microwave techniques.

ERTS

The Earth Resources Technology Satellite was launched by the National Aeronautics and Space Administration in July 1972 to make repetitive, synoptic spectral measurements of the earth. The satellite was placed in an orbit such that nearly every point on the globe was visible to the satellite during an eighteen day cycle. The satellite system includes a four-band multispectral scanner imaging system which provides spectral images of the earth in green (.5-.6 μ), red (.6-.7 μ), and two reflective (.7-.8 and .8-1.1 μ) infrared channels. These images are relayed to ground stations and are processed for wide dissemination in both image and digital computer compatible tape formats. For the first time, good synoptic and repetitive data are economically available routinely to a broad segment of the country. With the life span of ERTS-1 running short, the prospects for the launch of another ERTS series satellite become increasingly more promising. Apparently, this kind

of data, which holds the key to synoptic regional management of resources, is becoming increasingly and consistently available. It is also apparent from the amount of information which comprises an ERTS scene (2 million acre-sized picture elements) that traditional image interpretive techniques are inadequate and digital computer processing is the only viable method for data analysis.

Analysis Systems

Extensive analysis techniques have been developed for handling various forms of imagery. The large initial investment necessary to begin analyzing the more interesting problems offered by real data, as opposed to mathematical abstractions, accounts for the domination of this area of research by a few relatively large and well funded institutions. The analysis efforts have been concentrated largely in two major areas: (1) the extraction and enhancement of monochromatic images, and (2) the assignment of specific categories according to multispectral signatures. Image analysis and enhancement techniques have been pursued largely by the Image Processing Laboratory at the Jet Propulsion Laboratory [19] and also at the University of Southern California [20]. Computer classification from multispectral features has been introduced primarily at Purdue University [21], University of

Michigan [22], and University of Kansas [23]. Combinations of these techniques have been implemented on the NASA/JSC Digital Image Analysis System (DIAS) [24]. Software development in these areas has also taken place at the University of Maryland [7] and the University of Pennsylvania [25].

Recently, considerable emphasis has been directed at applying the techniques of computer graphics to the general areas of urban analysis systems, natural resource information systems and systems that concentrate on environmental problems [26]. Again the field is, to a great extent, dominated by the large facility, the computer-oriented designer and the sample solution, as opposed to an operational system.

Although great strides have been made in the acquisition, storage and automatic analysis of spatial and temporal data, the field is dominated by large computing machines and organizations, and intricate software systems. Yet according to the Rouse and Schell model [17], the benefit of the overall system must be measured by its effective use. Phillips [26] has observed,

The most formidable obstacle to more effective use of computers in solving urban and environmental problems is the lack of effective "interfaces" between the user and the programs he uses. The interface problem is especially acute because the user is not computer knowledgeable...there has

been almost no progress either in presenting data in a form useful for problem solving, or in providing functions to help transform data into useful information.

Management Actions

The resource management of large regional areas utilizing aerial imagery is quickly coming of age. Until recently, the technical capability of complex analysis of images was vested in only the largest groups which were able to afford the necessary computational equipment. This was complicated by the scarcity of the required timely images. With the advent of ERTS satellites and low cost minicomputer systems, the initial cost has become more realistic. Low resolution digital image displays capable of color enhancement are also becoming available at reasonable rates. The hardware components required to support regional management of resources seem readily available at this time. It is only the software aspects, the distillation of image analysis techniques into an operationally usable system that is not as yet fully developed. Great numbers of analysis software packages have been written to address many aspects of the image enhancement and analysis mode, but all of these have been written requiring the support of great computing centers and learned practitioners of the art of computer processing of image data.

It is important to remember that the utility of the whole remote sensing analysis system can be judged solely by the results of a user who is not computer knowledgeable. Nor can the user be generally expected to have an in-depth working knowledge of the physical characteristics of the measured energy, the engineering characteristics of the sensor systems, or the mathematical subtleties of the analysis systems commonly utilized in the remote sensing field. His expertise must be considered to be confined to the activity which he is best capable of performing, the analysis and planning related to urban, natural resource and environmental problem-solving.

Graphic and cartographic methods for analysis of spatial data are required by the resource manager. The planner is concerned with extrapolating land usage and its associated environmental impact. Analysis of spatial comparisons is accomplished by a conceptual overlay of a variety of thematic maps which depict regions of arbitrary shape where attributes are homogeneously distributed. Variables are retrieved and assigned relative weightings corresponding to their desirability in the environmental design. Combining the weighted variables for a specific location in a functional form provides a value of the desirability of locations in terms of the environmental design. Other information of interest also includes the

determination of the areal extent or number of certain environmental or resource events, and data presentation is most usable in cartographic or statistical display format.

CHAPTER III

INFORMATION SYSTEM DEFINITION AND SCOPE

Introduction

The application of remote sensor data to socio-economic problems is constrained by the rate of information flow across the data analyst-user interface. This rate of flow is impeded by the lack of awareness and understanding of each of these groups with the others' capabilities and needs. A link is missing between the survey and analysis techniques of the remote sensing community and users.

Most current analysis systems are not developed to a fully operational state. Software still requires an analyst who fully understands the intricacy of its techniques and of complicated execution procedures, and processing systems are either highly complex, general purpose digital computers with equally complex operating procedures, or simple devices which rely primarily on heuristic analyses. Computer analysis techniques are fully operational only when they are developed to the point where the analysis steps are specified methodically and catalogued with use and parameter explanations. Complex data analysis techniques are required in a form compatible with users' equally sophisticated, but solution-directed techniques. This void has been noted by an

applications panel convened by NASA [27]:

The users comprising this application panel, as a result of their discussions and interaction with the different panel experts in the sensor technology, concluded that their current applications capabilities for sensor data lag behind the sensor technology. This is to say that these representative users acknowledged that their particular communities must for the most part improve their capability (in both numbers and sophistication) to assimilate, process, interpret, and apply the relevant sensor data which is essential to their programs.

It follows therefore, that users, for the near term, must develop their own applications capability. This is necessary before they can specify incremental additions to existing systems or the development of entirely new systems based on the applications needs. This does not imply, as a corollary, that research and development in the sensors field can mark time until the users catch up, but rather that additional R&D efforts for the near term must be directed toward enhancing existing systems and improving the present quality of data reduction and processing.

Personnel concerned with the interpretation and action on resource items available through an information system are non-technical, from a computer analysis point of view. For the most part, these people are ecologists, biologists and managerial personnel with relatively little or no experience in computer operations, computer programming, or the technologies involved in remote sensing data analyses. It is these individuals to whom this system development research is being directed, for these

management functions for the most part have not been included in the specification of earth resources analysis systems.

System Design Criteria

Certain general characteristics are desirable in a resource information management system, especially in light of the expected system users. A specific system requirement is a simplistic user interface which is interactive and self-explanatory in nature. A suitable dialogue to clearly state and explain the system analysis options is important, as well as the availability of ample error diagnostics and recovery procedures. It would appear that a system which implements the analysis capabilities required for data interaction, classification and enhancement is desirable. For the most part, analysis procedures with specified design parameters are required. A methodology for the storage retrieval and handling of data, which is efficient, flexible and versatile, for the handling of source information forms the basic foundation of a resource information management and analysis system.

For the most part, the general method for storage, retrieval and handling of sensor data currently used is on a per pixel (picture element) basis. This is convenient from a technologist's point of view for it is in this

format that the sensor data are available and it is in this format that image display is most easily handled. Although the per pixel basis is convenient and effective for the handling and display of data, this method falls short in the handling of the information derived from the interpreted image, whether it be manually or automatically interpreted.

Resource information is obtainable by the interpretation of the spectral distribution of energy reflected from the individual elements which comprise the imaged scene. This information is spatial and temporal in nature and available from varied sources. A data format amenable to the handling and comparison of multiple data sets which distinguishes between the point, line and area aspects is required by an information handling system. Additional capabilities for data transformation and data coordinate scale and location changes are appropriate. Suitable output product displays which relate directly to a decision making function of a user agency or which might supplement the manager's intuitive feel about a given situation are also important.

Data Base Representation

Information Types

An operational resource information system translates data from imagery, surveys and administration records and relates the data to a common geographic base. In a resource information base, integrity between three basic information storage and retrieval formats is desirable. These formats include information stored as: 1) point information; 2) lineal information, and; 3) areal information.

Point information is that information which relates to a specific coordinate location or similarly, that coordinate point which locates the information. Standard sensor reflectance data are of this variety. A measurement of the spectral reflectance from a single ground resolution element is associated with the coordinate of that ground point. Also, any interpretation of the spectral data is associated with the point coordinate. Another association of this type might be the location of a specific water tower, for example. Here again is the interpretation/coordinate association for a single ground point.

Lineal information is characterized by the association of a single entity or data value for a number of ground resolution elements connected in a lineal fashion.

Examples of this type of association would include roads and area boundaries. This representation allows the aggregation of point information of the same value which may be represented by a larger more precise entity.

Areal information is characterized by the association of data in an areal manner. Where adjacent ground cells have the same information associated with them, they may be accumulated under a single label and associated with the information. An example of this type of association might be all the resolution elements of a single cropped field (e.g. a cotton field) or might relate to a particular political association such as an area zoned for commercial development.

Association of information under labeled aggregations of the type described allows for handling of large numbers of point/information associations within a single aggregated unit. Thus, it not only reduces the storage required for the data, but allows for a more flexible interpretation and transformation of large volumes of data. Information stored under these separate types of labels is also easier to edit, expand and delete. The major problem in this labeled representation method is how to store, retrieve and analyze data with such widely varying data formats, and the definition of the specific formats of each of these data types.

Raster Format and Display

A system of small area units based on geographic coordinates provides a flexible means of collecting, storing, performing computer enhancement and decision operations, and display of collected data. In this basic raster format, values are associated with a coordinate which is implied by a position or location in an array or string. Imagery is either collected or digitized generally by some form of raster scan. Many image sensors in the current earth resources survey inventory provide digitized data in this format. Digitization of photographic data also is accomplished in this format. Computer manipulation of the image elements as pixels in a random access stored array is extremely convenient, and image display of the image data by printer graphics or on color television units is greatly facilitated. It would seem that the availability of an image as picture element values in a random access array is the most suitable machine format.

Only through image interpretation analysis (either interactive, manual or automatic) do the sensor measurements become resource information, that is, knowledge required by an interpreter, analyst or manager which can be utilized in making a decision. The important characteristic is the location of a feature on the ground and its association with other similar, dissimilar or

aggregated features. In the sequential format, it is difficult to isolate individual units without reconstructing a large portion of the data set in a random access storage medium. This sequential digitization and storage method for all practical purposes precludes scale and coordinate rotations of the data for congruencing with similar data sets from the area. The storage method also is cumbersome in that all the data must be represented by individual data cells of the smallest integral size desired.

Sensor measurements in and of themselves are useful only in as much as they provide additional knowledge about the location, relationship, condition, existence, etc. of the resource categories of concern. The termination of the digital analysis chain with the production of a raster format product, whether image or classified image, is unfortunate. It constrains the resource manager to the role of image interpreter, since the terminal product of almost every image analysis system is either the raster scan array stored on magnetic tape, magnetic disc, printed on a photograph or displayed on a CRT.

Boundary Representation

Upon interpretation, the significance of individual picture element values is supplanted by the category assigned to their aggregation. A flexible, efficient methodology is important in the handling of these element aggregations. It is also desirable that this methodology be sufficiently flexible to handle information available from other sources relating information to spatial aggregations. Such a methodology can be built around a data format which aggregates ground points by defining the boundary which encloses or links them. This format can handle point, lineal and areal types of data in an efficient and flexible manner.

A suitable data handling methodology for an information management system utilizes the boundary format, associating category values with the explicit coordinates of the boundary which links them. Utilization of parametric functions which describe the boundary can minimize the amount of data stored. For example, if a piecewise linear boundary is assumed, only the coordinates of the line segments are saved as the information. In most cases, piecewise linear fits are adequate for data description, although quadratic or other parameter functions can also be utilized.

Edge-tracing algorithms can automatically convert

raster information such as classification-interpreted images into the boundary format. Interactive conversion of interpreted information into the boundary format provides the simplest approach to the format conversion. Figure III-1 shows an area of Brazos County as imaged by the Earth Resources Technology Satellite and displayed by means of a line printer representation. The line printer image representation is typical of the raster format which is generally the analysis product. On the image product are shown interpretations of resource features which might be important to the area. These interpretations are representative of the point, line, and area information aggregations which are important. Piecewise linear boundaries are also indicated.

The boundary format greatly reduces the amount of storage required for information maintenance and provides a flexibility unparalled by raster storage. As significant as the data reduction is the ease with which coordinate transformation may be enacted. With explicit coordinates, the transformations of translation, rotation and scaling become simple mathematical mappings involving no actual change in the elemental location of stored data. This format also facilitates the congruencing of data sets unlike either in scale or orientation. This boundary format representation is flexible, efficient, data-reducing

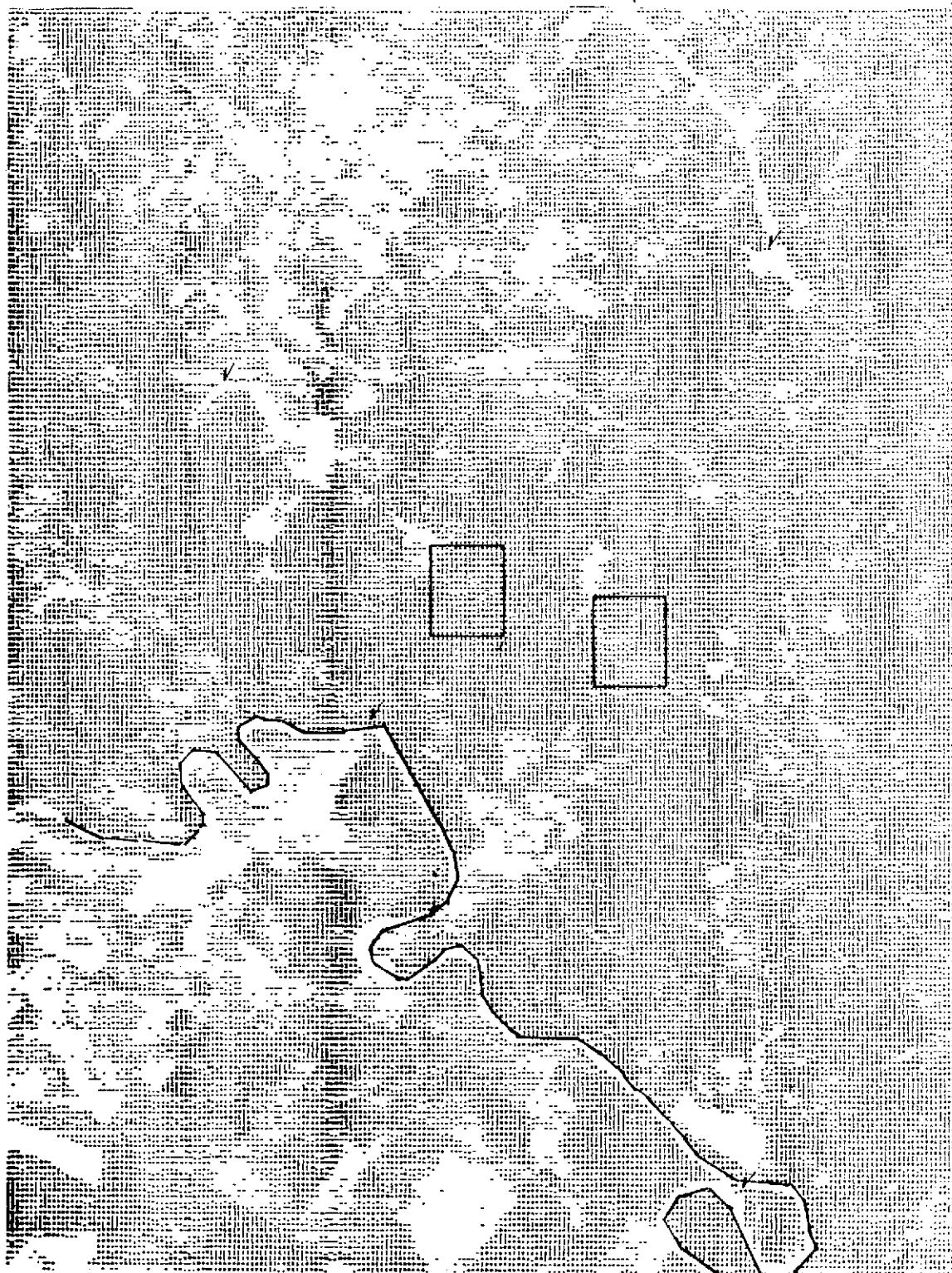


Figure III-1. Computer Processed Grey-scale
Image of Brazos County

ORIGINAL PAGE IS
OF POOR QUALITY

and convenient.

The techniques appropriate for the conversion and handling of resource information in the boundary format are not particularly novel, however, the presentation of an information storage and retrieval format which provides a convenient and efficient methodology for transferring, handling, comparing and analyzing interpreted image sensor derived information is a significant step forward in the application of remote sensor data to the problems of resource management.

Transformations and Data Handling

Coordinate Transformations

In a raster storage mode, the typical format for image remote sensor data, values are associated with a coordinate which is implied by the position of the value in an array or string. Coordinate transformation requires the repositioning of the value element in the storage array. Three basic coordinate transformations are required for the handling of data. These are translation, rotation and scaling, which are extremely important when comparing data sets which differ in the time acquired sensor resolution, or format. Scaling and coordinate translation are not mechanically difficult for raster stored data, however, rotation is an extremely cumbersome

and time-consuming transformation. This is particularly true because coordinates are implied by the position at which the data are stored in an array.

The coordinate transformations are no longer cumbersome procedures for data stored in the described boundary format. In this format the spatial coordinates of the data values are explicit and the storage of the data values is not constrained to a fixed location. Coordinate transformation of data in a boundary format relies solely on the mapping of the coordinate values associated with the information. With the explicit coordinate values, the transformation reverts simply to the application of linear coordinate transformation equations. The incorporation of an interactive coordinate transformation capability in an information management system is particularly important in congruencing data sets from differing origins or in locating information on imagery by overlay of familiar features.

Data Transformation

Resource inventory and control decisions are based particularly on observing the juxtaposition of favorable or unfavorable aspects of the environment. Under the "right" set of conditions, for example, a wildlife habitat may be suitable. A change in one of those condi-

tions may degrade the habitat to a marginal or even hostile one for the wildlife species under consideration. The kinds of information which are used, for example in determining a suitable habitat, are available from a variety of sources and may be presented in a variety of ways. The resource information management system which is used to aid in the making of environmental decisions requires the capability for logically comparing and combining decision elements into a single environmental quality value. Suitable mathematical and logical operations can be interactively defined and utilized by the system operator. These operations include logical operations, conditional operators, and mathematical operators.

Format Transformation

An important aspect of a resource information management system is the ability to convert data from one format to another. The boundary coordinate format described previously is suitable for an external format and a raster format is appropriate for display and image operations by an information management and analysis system. Conversion from an external boundary format to an internal raster format requires algorithms to expand point, line and area information. Interactive procedures

provide the simplest method for reversion from raster format back into a boundary format with the system operator providing the coordinate points defining the boundary and the information values associated with data entry.

Analysis Operations

The literature [6] describing the analysis of image data is resplendent with glowing reports of algorithms for spectral classification and image enhancement. These algorithms are the "answer" to the application of multi-spectral image sensing and provide meaningful inputs to the functional block labeled "user". These descriptions are just as conspicuously void of any description of how the image-derived data products can be applied and of any description of how the image formatted interpretations are available to the "user", unless he reverts to reproducing all the enhancement and classification results to display them on his own system. These descriptions, however, provide a comprehensive treatment of the various image analysis techniques which have been utilized to advantage in applying remote sensor data, and provide an adequate basis upon which algorithms may be selected and implemented in a resource information management and analysis system. Two fundamental analysis techniques required by the system are classification, and image enhancement.

Computer-Aided Classification

Computer-aided classification is an algorithmic procedure for the computation and comparison of similarity measures between sensor measurements and the assignment of similar measurements to identified classes. A class or elemental signature is that set of measurements, characteristics or features which have been obtained from that class, or elements which describe the best knowledge obtainable by the sensor or sensors used to measure the object. The term "signature" is used to imply a unique association between the measurements and class elements, although the uniqueness of the association is not always supported. The primary assumption upon which classification is based is the assumption that measurements of dissimilar objects are not alike. The algorithm attempts to measure the similarities for classification purpose.

There are two general approaches to the classification problem. These are supervised classification and unsupervised classification. Supervised classification is characterized by the a priori assignment of data elements to classes and the calculation from these assignments of typical class signatures against which unknown elements may be compared. The actual training procedure depends upon the actual classification algorithm utilized.

The most commonly used classification algorithm is

based upon a maximum likelihood criterion. In this procedure, the form of probability density functions on the classes involved is assumed or determined. The measure of similarity between a data vector and a class is the a posteriori probability of the data point given the class which is most likely to have produced the point. The assumption is generally made that the distributions are multivariate Gaussian and the class sample means and covariances are used as estimators of the parameters of that distribution. This is the classification algorithm which forms the basis for the LARSYS series of classification algorithms developed at the Laboratory for Applications of Remote Sensing at Purdue University [28].

Unsupervised classification or clustering relies on the aggregation of similar data points and the a posteriori assignment of classes to these aggregates. The procedure may be single pass where data points are examined only once, or multiple pass where data points are examined several times in an attempt to improve upon the previously determined aggregations. Similarity measures are calculated between the data points and the aggregations or clusters and the point is assumed to belong to the most similar cluster if the similarity measure is below some predetermined threshold. For this case, the cluster parameters are improved by the new data point and the

procedure continues. If cluster dispersion exceeds another threshold, the clusters may be split into two new clusters. If the data point is determined to be dissimilar from all defined clusters, a new cluster is formed by the data point. Differences in clustering algorithms are primarily due to differences in defined measures of similarity and in the procedures by which new clusters are formed or old clusters split. An attempt is made in some clustering algorithms to link clusters at the end of the procedure so that cluster shapes and sizes are not completely dependent upon the similarity measures and the thresholds chosen. Some versions of the LARSYS procedures use a combination of the two techniques by clustering within training set collections to improve upon the points selected for training and to give better estimates of the actual class signature parameters.

Image Processing

Certain image processing functions are desirable in an analysis system. These processing operations are utilized to improve the interpretability of the imagery. Processing operations may be classed according to their mode of application: neighborhood, non-neighborhood and transformation operations. Neighborhood operations transform a data point according to the values of the

data points which surround it, that is the transformation is coordinate location dependent. An example of this class of operation is local averaging where a point is replaced by the average of the neighborhood to which it belongs. Non-neighborhood functions transform the data irrespective of its coordinate value or location. These operations involve scale changes in the image function such as grey-scale reversal or density slicing. Image transforms such as the Hadamard and Fourier transforms, deal with the mathematical mapping of the image space into another space.

Several imaging processing functions are desirable in an image analysis system. Of particular importance are smoothing, edge enhancement and grey-scale transformation operations. Smoothing and edge-enhancement may be accomplished either by neighborhood operations or by transformation. The simplest and most straightforward approach is the use of neighborhood operations.

Communications

Output Products

Information management and analysis system output products are to a great extent limited by the types of peripheral output equipment available. Computer systems commonly include printer output and a CRT output capability

is desirable. General categories of output then are usually constrained to printer tabulation and graphics, and CRT image presentations. Other output to sequential file systems is appropriate for created or modified data files.

Operator Communication

The importance of an interacting man in an information system is emphasized by Martin [29].

...man must become the prime focus of system design. The computer is there to serve him, to obtain information for him, and to help him do his job. The ease with which he communicates with it will determine the extent to which he uses it. Whether or not he uses it powerfully will depend upon the man-machine language available to him and how well he is able to understand it. To be effective, systems will have to be designed from the outside in. The terminal or casual operator instead of being a peripheral consideration will become the tail that wags the whole dog.

Design of the dialogue for operator communications requires consideration of the level of training of the expected operators of the system. Although these individuals are intelligent, their exposure to computer languages and processing operations is generally limited. The operators for the most part are casual operators as opposed to operators dedicated to system operation. These casual operators need to use an information and analysis

system at some irregular frequency and for the most part cannot be expected to remember complex mnemonic entry nor the full sequence of operations or capabilities of the system.

In view of these operator considerations, the computer-initiated dialogue is preferred, refreshing the operator on the full capability of the system whenever required. Certain functions of the system require entry of programming-like statements, in particular for coordinate transformation or for data comparison and decision, however, this interaction should be kept simple and to a minimum.

Another procedure for communication with the operator described is menu selection. Fixed responses are expected in particular for specification of the processing techniques. Display of the various options in a menu format with a method for selection of the desired technique offers a reasonable approach to operator communication. Flexible but simple and informative operator communications are necessary for an effective interactive system.

Error Diagnostics

Terminal operators usually make errors. It is essential to protect the computer system and its data from these errors, as far as is possible, and to do this

without antagonizing the operators. A factor that is strongly in favor of real-time systems is that in an appropriately designed dialogue, most operator errors can be caught in real-time. The mistake or discrepancy is detected and corrected on the spot. The effectiveness of error detection is highly dependent on the design of the man-machine dialogue.

Psychological considerations are important in the design of a man-machine dialogue. It is desirable to steer a course between boredom and bewilderment. Studies of the human learning process have shown that positive response to correct actions and admonition of incorrect actions within seconds give by far the best reinforcement [30]. The conscientious operator learns from the response to his incorrect actions. The response, however, need not be overly abrupt: a split second error response amid thought is jarring and rude. The error response suitably is delayed until the operator is permitted to finish his thought. One danger in giving real-time error responses is a tendency for an operator to become careless expecting the machine to catch all of his mistakes.

The requirement for error detection and diagnostic response is apparent. However, a dialogue critically examined in light of both motivational and technical considerations is more effective.

IRIMAS--An Interactive Resource Information Management and Analysis System

The specifications outlined above are embodied in an interactive resource information and analysis system--IRIMAS developed on a Texas Instruments 980A minicomputer system. This is an interactive software system configured for simplistic analysis of resource information derived from remote sensor data interpretation. It is also suited for inclusion of resource information from other information sources. The salient features of the system include a data handling system for the storage, retrieval and analysis of resource information in the boundary format described. The data are encoded for extended storage in data strings where the boundaries are defined by piecewise linear curve fits to the coordinates. System flexibility is maintained through the utilization of syntax definitions and syntax interpretation algorithms.

Syntax algorithms are utilized for: (1) conversion of the external information character string into an internal format suitable for further analysis; (2) the interactive definition of data-and coordinate-transformations in programming-like statements, and; (3) the interactive operation interface to the information system for the specification of system analysis functions.

Algorithms for the analysis of remote sensing and

information data are implemented to include the following capabilities: (1) CRT and printer map display, (2) maximum likelihood classification, (3) image enhancement with smoothing, edge enhancement, and single pass clustering, and (4) computation of data statistics.

An interactive operator interface is implemented to allow simplistic operation of IRIMAS by specification of function procedures for standardized data analysis.

In this chapter are discussed the general requirements for a computer system to perform interactive analysis of resource information and image interpretation. A mini-computer system is developed using these general guidelines and specifications to address the problem of interfacing image data available from remote sensing sources to the resource management function. The specific components and implementation of this computer system, IRIMAS are subsequently described in following chapters.

CHAPTER IV

IRIMAS DATA HANDLING AND TRANSFORMATION COMPONENTS

The general design goals for an interactive information and analysis system which aids information management and image interpretation by a resource planning function are described in previous chapters. An Interactive Resource Information Management and Analysis System, IRIMAS, developed with these design goals is implemented on a Texas Instruments 980A minicomputer system. Although the implementation is restricted to use on the TI 980A, the techniques employed are applicable to any general purpose computer system with suitable display peripherals. Subsequent discussion describes the techniques applied in IRIMAS and the specific implementation employed.

The more significant features of the general information management system described and of IRIMAS in particular are concerned with the encoding, decoding and handling of point, line and area information. In IRIMAS, the boundary format previously discussed is used for information storage and retrieval. The flexibility required by the system is met through the use of syntactic definitions and algorithms for syntax interpretation. Configured in this way, modification of the interactive input and output procedures and changes in the specific format utilized for information representation mostly involve syntax definition

change. These changes are localized to specific sub-routines which are easily modified; the major logic and processing components of the system remain unchanged.

In this chapter are discussed the background required for the understanding of formal language procedures and the software devices required for interpretation. The specific components of IRIMAS required for interactive transformation of data and coordinates are discussed, as well as those components required for interpretation of the information storage format. System algorithms and software components used in expanding the boundary coordinate information representation into the raster or grid data arrays is also covered.

Formal Languages

Formal language theory has developed an abstract terminology for the modeling and representation of information structures which can be represented by a set of primitive elements together with a set of rules for combining those primitives. The formalization of languages and their representation has allowed the development of general procedures for the recognition of sentences in a suitably defined language without undue concern over a specific language definition. A subclass of all languages is the set of languages defined by phrase-structured

grammars which have been widely used in the definition of formal computer programming languages such as FORTRAN, ALGOL, etc. The translation of these programming languages is done mechanically by compiler software which implements the general syntax-driven parsing procedures. These language techniques have been used to define a data structure for the data maintained by the IRIMAS information system, and for the definition of operator language and command, and interpretation.

Language Terminology

The terminology common to formal language specification is defined in references [31] and [32]. An alphabet is any finite set of symbols used to converse in a language. In the English language, this would correspond to the well-known 26 letter alphabet plus those symbols or punctuation which might also be appropriate. A sentence is any string (or sequence) of finite length, composed of symbols from the alphabet. A language defines a set of sentences over an alphabet domain. In the formalization of a language, the need to recognize what alphabet strings belong to the language, where the number of sentences in a language is infinite, becomes a problem. For certain languages, the sentences of the language can be generated by a procedure which is formalized by the concept of a

grammar. A grammar is the formal definition of those sentences which generate a language. The structural definition of the language in terms of a grammar is called a syntax of the language.

These terms and their relationship might best be demonstrated by a sentence from the English language. As an example take the sentence, "The black cat ran away." In terms of English grammar, this sentence is composed of a noun phrase, "the black cat", and a verb phrase "ran away". The noun phrase consists of the modifiers "the" and "black", and the noun "cat", while the verb phrase consists of the verb "ran" and the adverb "away". The grammar for this sentence is defined in terms of (1) syntactic categories from which strings of words could be derived; (2) variables or non-terminals; (3) terminals or words; and (4) productions, which are the relations that exist between various strings of variables and terminals. A grammar also specifies a non-terminal "sentence" or "start" symbol which generates exactly those strings or terminals that are in a language.

The presentation of the example sentence in terms of a formal grammar and its notation is as follows:

```
<sentence>::=<noun phrase><verb phrase>
<noun phrase>::=<modifier><noun phrase>|<noun>
<modifier>::=The|Black
```

```

<noun>::=Cat
<verb phrase>::=<verb><adverb>|<verb>
<verb>::=<ran>
<adverb>::=<away>

```

The grammar consists of non-terminals or variables such as <noun phrase>, <noun>, <adjective>; terminals or words such as "the," "cat," "black," and productions are those relations that exist between various strings of variables and terminals. The symbols "::=" stands for "is defined to be", "|" stands for "or" and the angular brackets are used to enclose each name of a phrase or variable. The formal grammar also requires a starting symbol which in this case is the variable <sentence>.

An additional method for representing a grammar is in terms of a syntax table. A syntax table is the tabular form extensively utilized by the information syntax algorithms. A syntax table for the above sentence may be constructed as shown in Table IV-1.

Syntax Driver

A general purpose driver may be formulated to step through syntax tables to determine if a sentence belongs to the defined grammar. The driver will try to match a sentence like "The black cat ran away" to the syntax tables. In doing this, it must keep track of the current

TABLE IV-1
Example Syntax Table

sentence	equation equation end of table	noun phrase verb phrase
noun phrase	equation or equation equation end of table	noun adjective noun phrase
adjective	literal or literal end of table	The Black
noun	literal end of table	Cat
verb phrase	equation equation end of table	verb adverb
verb	literal end of table	Ran
adverb	literal end of table	Away

syntax position.

The sentence is held, internal to the computer, as a string of words. An input cursor is used by the driver to position itself in the sentence. When processing begins, the cursor will point to the first word in the sentence and as processing continues, the input cursor will advance.

As the sentence is processed, sentence words are matched against items in the syntax tables. The driver will keep track of its current syntax table position as it sequences through the table. With each syntax table position is associated a syntax position type and a position value. The syntax position type designation indicates whether the position value is a variable (equation), a terminal value (literal), or the type dictates a table delimiting function (or, end of table). Two additional types of syntax position are also utilized. The position type "repeat" indicates that at this point in the syntax table the non-terminal value is considered to be repeated zero to an arbitrary number of times. Syntax position type "subroutine" indicates that the position value is to be output to the interpreting subroutine as an operation code. The position of an output string maintained by the interpreting subroutine is also available to the driver.

The driver maintains push-down stack storage of the

input position, the syntax position and the output position. It keeps a series of these pointers in the stack as it works its way through the table for each equation. Then, when the driver is ready to return through the equations, it has a memory of its path. Having taken a wrong path through the syntax equations, the driver also has a memory of input and output cursors and thus can restore the input and output cursors from the stack to try an alternate path.

In Figure IV-1 is shown a flow diagram of a general purpose syntax interpreting algorithm. Syntax interpretation is handled by four components of the syntax driver corresponding to the syntax table position types:

(1) equation/repeat, (2) literal, (3) or/end of table, and (4) subroutine call. As the type of current syntax position is determined the appropriate software component is utilized to further process the input string. The syntax driver outputs operation codes by way of the subroutine call and halts with a "valid syntax" decision if the input string exactly matches the syntax table, or an "invalid syntax" decision if the two do not match. For a detailed description of the syntax driven algorithm see Gauthier and Ponto [34].

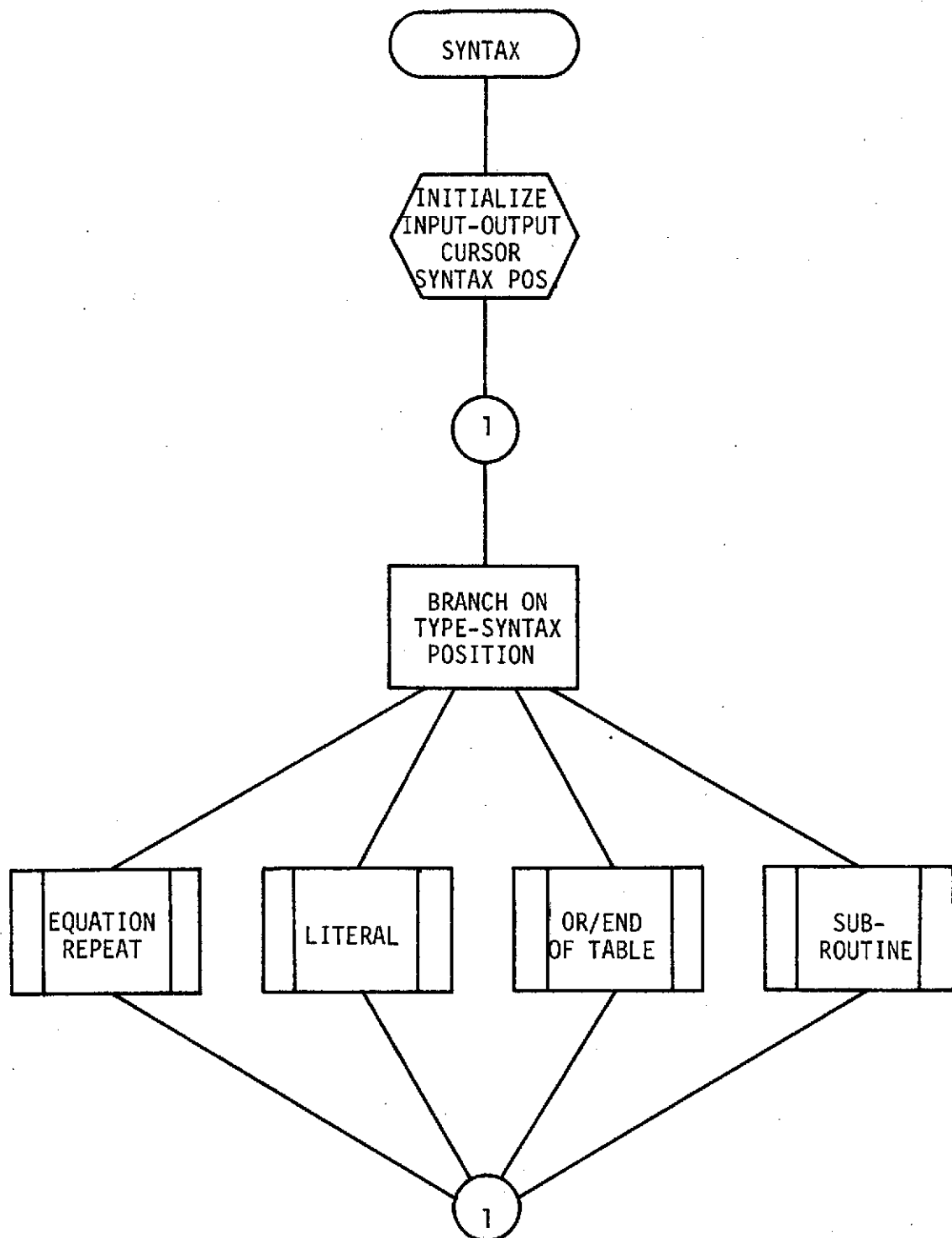


Figure IV-1. Logical Flow Diagram for Syntax Interpretation

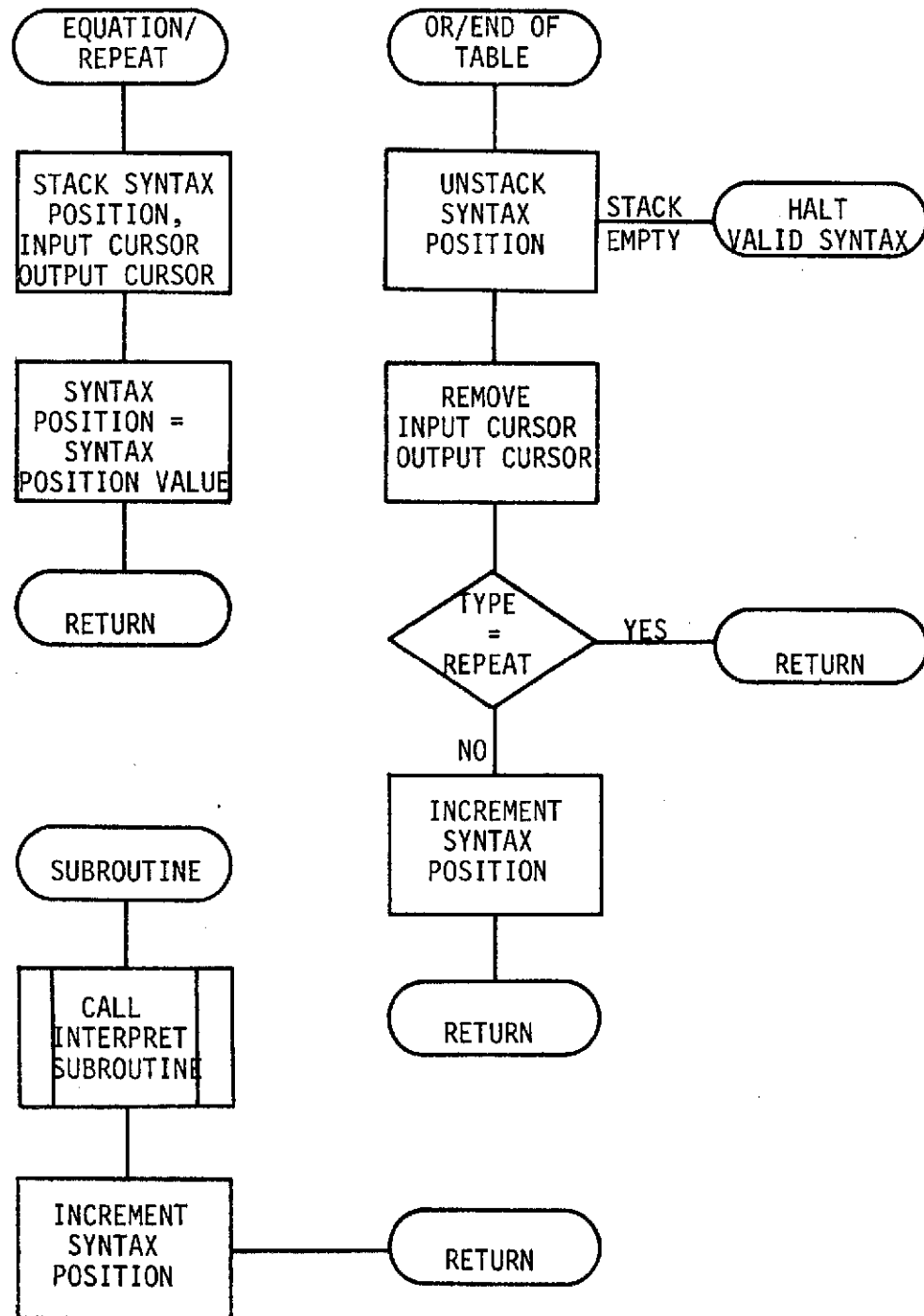


Figure IV-1. Logical Flow Diagram for Syntax Interpretation (cont.)

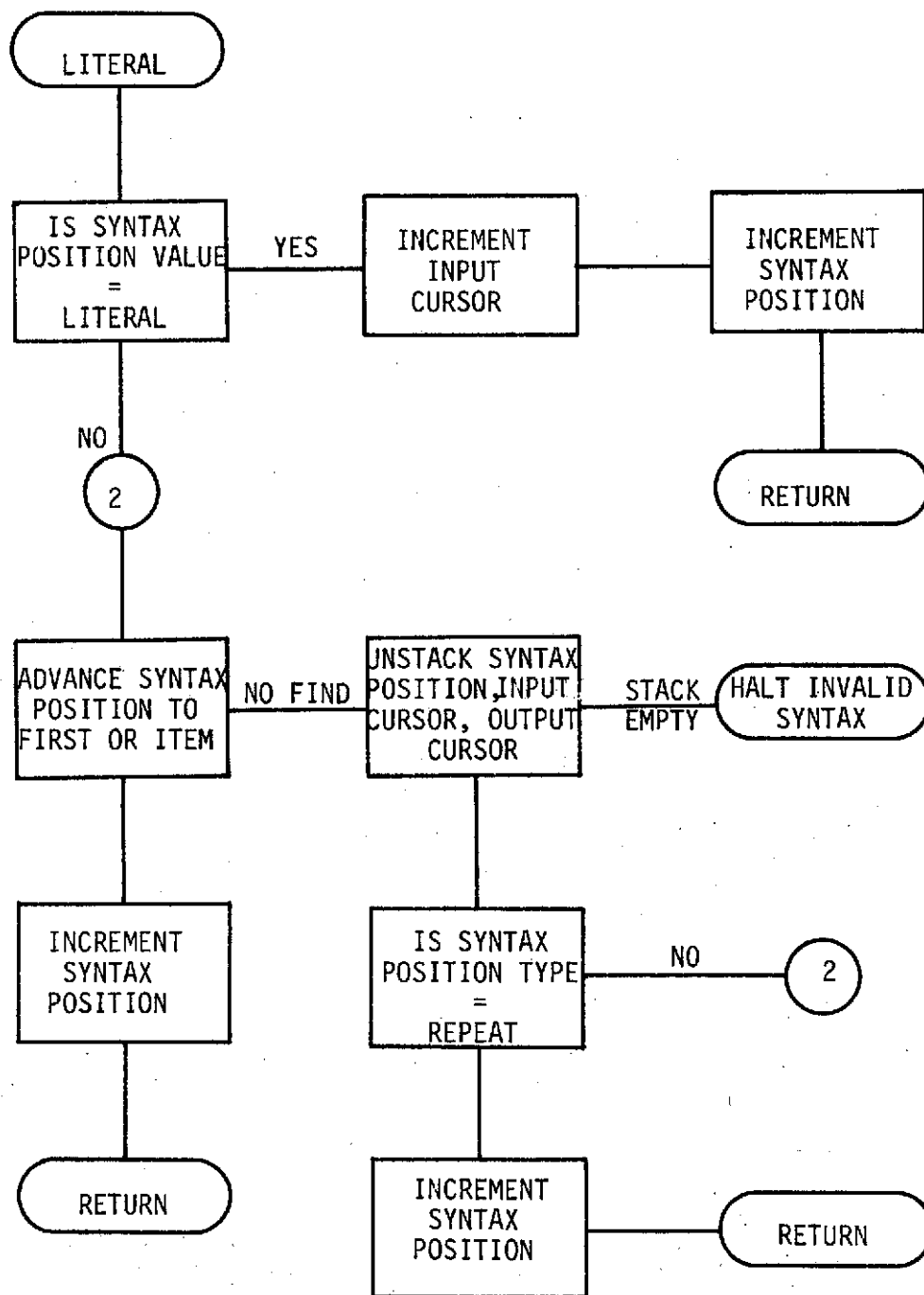


Figure IV-1. Logical Flow Diagram for Syntax Interpretation (cont.)

IRIMAS Information Syntax

A flexible resource information data system must handle data files containing point, lineal and areal information entries. These entries must be referenced to a common geographical coordinate base and must be flexible enough in their definition to include a broad variety of numerical, literal and logical information. A data base with such varied and flexible entries can best be built and interpreted if syntactic representation and parsing algorithms are used. This notational representation facilitates the definition of a hierarchial structure and allows mechanical interpretation or parsing of an arbitrarily defined language within a specific class of grammars.

The format of the IRIMAS data base is defined in terms of character strings which are collected into phrases and sentences similar to a spoken language. Each sentence has a different functional purpose in the precise definition of the information type and its value. Sentences in this data base definition have the following purposes:

(1) data set identification; (2) attribute definition; and (3) data definition. Data definition statements provide point, lineal and areal data entry. Additionally, a grid or raster format has also been included as a part of the information system format.

The data set is of a sequential nature and the first

item is a data set identification and description label. Items included in this label are the file name, geographic coordinate references and a narrative description of the data file to include such items as origin and previous processing, together with other descriptive material. Optional data set inclusions following the description label are "sentences" describing and defining the list of attributes associated with the data file. Numerical codes used to reference lengthy data cell descriptions would be defined in these data set entries. The third type of data set entries are the sentences defining the data. These entries would include a data entry label which describes the entry name and the type of entry (point, lineal, areal, or grid), a coordinate definition referenced to the coordinate definition included as part of the data set label, and the values of attributes applying to this specific data entry. These attribute values could either be quantitative or descriptive. Other items included in the data set would be delimiting punctuation between sentence entry items.

Block Label

The first entry in a block of data entries is a Block Label which contains identification and descriptive information about the entries to follow. This specific

entry is free form, where specific elements are positionally located. That is, items in the entry must follow a fixed sequence, but their lengths and absolute location in the data record are unconstrained syntactically. All characters in the entry are alphanumeric. An entry identifier, the first byte of the entry, is arbitrarily assigned the ASCII value of "\$". Thus, the first item in the sentence is the sentence identifier. Subsequent items are separated by either spaces or commas. The second item to be included is the data set name. This is the label by which the block is uniquely identified and logs may be kept relating to the origin of the information contained in the data set. The third and fourth entry items are the geographical coordinate reference for the block. Their exact natures are undefined here but longitude-latitude might be a typical entry item.

A narrative description of the data set may follow the coordinate values. This description would include any information pertinent to the data set which would be necessary at analysis time. The last item in the entry is a delimiter code arbitrarily selected to be an ASCII "/". This delimiter is used to separate sentence entries into the data set much like a period in a regular language.

Attribute Definition

The second type of entry, the Attribute Definition, allowed in the data set provides for the more precise definition of data attributes. This entry has no operational significance and is purely descriptive. Its purpose is to provide a more complete description of the values assigned to an attribute variable. Items within this entry are delimited by colons (:) and commas. Spaces are recognized only as valid characters. The first item in this entry, like all other entries, is a defining operation code, selected in this case to be an ASCII "&". The second item is the attribute name to be described. This name is separated from the subsequent descriptive information by a colon. The additional items in the entry are used to describe values assigned to the attribute in later data entries (e.g. 0 = no classification selected, 1 = cotton). These descriptive items are delimited by commas and the attribute definition entry is terminated by the delimiter "/".

Data Entries

The third type of entry allowed in the data set is the Data Definition statement. This sentence consists of several phrases each identified by a characteristic operation code and each phrase includes several items. A data

label phrase is used to initiate a data definition entry. This phrase includes an ASCII "*" operation code as the first item, which designates a standard type data entry. The next item is an alphanumeric "P", "L", "A", or "G" to define the data type. The letter "P" designates a point entry, "L" a lineal entry, "A" an areal entry and "G" a grid entry. The final item in the data label is the unique name associated with the entry. No delimiters are recognized in this phrase, which is terminated by the following phrase operation code. Once the data definition entry is opened by the data label phrase, the order of additional entry phrases is not restricted.

A second type of phrase for the data entry is the coordinate definition. This phrase begins with the operation code ASCII "%". A single occurrence of this phrase is sufficient for a point data entry; multiple coordinate entries, however, are required for both lineal and areal entries. Lineal entries are defined by the sequence of coordinates derived by approximating the curve with straight line segments. The coordinate entries are the endpoints on the curve of those lines. Areal entry coordinates are similarly derived only the curve closes on itself. The area defined is as that enclosed by the curve and the attributes entries refer to that area. Attributes of a lineal entry describe the curve.

The second entry in the phrase is the coordinate variable (either "X" or "Y") followed by the coordinate value. These items are followed by the second variable name and the coordinate value where applicable. In a differential mode, these additional items are optional. The omission of a coordinate variable definition in the differential mode is taken to mean no change in that variable value from the preceding coordinate entry.

The mode of the coordinate insertion is the fourth item in the phrase. An alphameric "A" designates absolute and an alphameric "D" designates the differential mode. Note that each data entry must have at least one absolute coordinate definition as the first coordinate definition.

Two types of data value-defining phrases are specified. The first beginning with an operation code "@" includes attribute values which are descriptive in nature. The second item in the phrase identifies the coordinate by its position in the block definition statement. Alphamerics are used for this identification. The letter "A" designates the first attribute "B" the second attribute, etc. This attribute designation is followed by the descriptive attribute value. For the quantitative attribute specification, the phrase begins with the operation code "#" followed by the attribute

designator as above. The next item in the phrase is the numeric string defining the value of the attribute. These phrases are terminated either by the beginning of another phrase within the entry or the entry delimiter "/".

The IRIMAS data entry statement also provides for gridded format data entry. Entry statements for gridded data utilize the coordinate entry format to specify the upper left-hand corner and the lower right-hand corner of the rectangular data block entered. The first attribute "A" is the number of 8-bit bytes in each picture element in the gridded array. After the data entry statement, the data are entered in a sequential manner with x-coordinate values changing most rapidly. The data entry statement terminates after all the specified data has been input.

Formal Syntax Definition

The statement definitions described above may be more formally represented in terms of syntactic definitions:

```

<entry> ::= <block label> | <attribute> | <data>
<block label> ::= $ <name> <block coord.> <description>
<attribute> ::= & <attribute name> / <description>
<data> ::= <data label> <coordinate> <data>
<name> ::= / <alphanumerics>
<block coordinates> ::= <longitude> <latitude>

```

```

<longitude>::=<degrees>:_<minutes>:_<seconds>
<latitude>::=<degrees>:_<minutes>:_<seconds>
<degrees>::=f<numeric>
<minutes>::=f<numeric>
<seconds>::=f<numeric>
<alphanumerics>::=f<alphameric>f<numerics>
<description>::=f<alphanumerics>
<data label>::=*<type><name>
<type>::= A|P|L|G
<coordinates>::= %<mode>f[<variable>f<numerics>]
<variables>::= X|Y
<data>::=<descriptive>|<quantitative>
<descriptive>::= @<attribute><alphanumerics>
<attribute>::=<alphameric>
<quantitative>::= #<attribute>f<numerics>
<alphameric>::= A|B|C|D...|Z
<numerics>::= +|_|1|2...|0

```

Machine Format

The data format specified for information storage and retrieval is inconvenient for manipulation and analysis by computer software. Consequently, syntax driver parsing algorithms are used by IRIMAS to convert the system data language into a convenient format for machine analysis. These parsing algorithms convert the syntax data into the

constituent elements of alphanumeric descriptions, coordinates, data type, attribute values, and also set attribute flags to indicate that specific attributes are related to descriptive text rather than quantitative value.

The data for each entry element are maintained in several arrays. The array containing the alphanumeric descriptors contains the entry element name as the first string in the array. Additional strings in the array are produced by those attribute entries which have descriptive values rather than quantitative values. The length of the string is stored as the first word-sized element in the string and the string is word-aligned. Coordinate data is stored as x-coordinate and y-coordinate pairs in a stack array. The first element in the array is the number of entries in the stack. The coordinate pairs are stored as absolute values referenced to the data set beginning coordinates rather than in the differential mode which is also permissible in the data entry format.

The attribute array contains several kinds of information. Attribute numbers are fixed and each attribute entry is either an attribute value for those attributes with quantitative values, or the position contains a pointer to the string in the alphanumeric descriptor array which corresponds to a description value of the attribute. In order that there can be no confusion as to whether an

entry is a value or a pointer, a bit is set to one in the first two words of the actual attribute array. The bits of the double word correspond to the attribute position in the array with the high order bit corresponding to the first specified attribute. The third coordinate in the array indicates the type of data entry with point, linear, area and grid entries corresponding to type values of 0-3, respectively. This in essence leaves the fourth position attribute, designated "A" in the data entry format.

Grid format entries from the data set utilize the coordinate array to store the coordinate extents of the grid input, and utilize the first attribute value to specify the number of bytes corresponding to each coordinate point entry. Specific quantitative values input in the grid format are maintained in an additional buffer area reserved for grid data. It is in this machine format that the data entered into the IRIMAS software system is further analyzed.

Data Syntax Conversion Subroutine

The IRIMAS system uses a conversion subroutine SUBR.1, together with the syntax driver algorithm to convert the data as interpreted by the driver into the machine data format. Conversion is controlled through a set of eight operation codes to the subroutine which are

embedded in the syntax tables defining the information system data format (see Table IV-2). Additionally, the current position of the input cursor is provided to the syntax driver to allow correction of data for improper traces through the syntax table.

Several buffer areas are maintained within the conversion subroutine software. Buffers for current x and y coordinate values (CV.X, CV.Y) are maintained as well as a buffer location (TEST) which is used to hold characters from the input string for comparison. A pointer to the output buffer containing data alphanumeric descriptions is maintained which points to the beginning of the character string associated with the current syntax variable being processed by the conversion subroutine. The output cursor generally points to the location of the last character in the output buffer. An additional buffer area (VALUE) is provided which holds the address pointer to the location of the current attribute or coordinate value being processed.

Conversion subroutine operation codes, their syntax implications and subroutine processing are described in Table IV-3. In Figure IV-2 is shown the logical flow chart for the conversion subroutine.

When the syntax-driven algorithm is utilized together with the appropriate syntax table and the conversion

TABLE IV-2
Information Syntax

Entry	Subroutine Repeat Equation Repeat End of Table	0 Delimit Label Cr. Atr.
Cr. Atr.	Equation Or Equation End of Table	Coordinate Attribute
Attribute	Equation Or Equation End of Table	Data-Alpha Data-Numeric
Label	Literal Equation Subroutine Equation Subroutine End of Table	* Type 1 Name Alpha
Type	Subroutine Literal Or Literal Or Literal Or Literal End of Table	3 A P L G
Name	Repeat End of Table	Alphanumeric
Alphanumeric	Equation Or Equation End of Table	Alpha Numeral

TABLE IV-2 (cont.)

Alpha	Subroutine	3
	Literal	A
	Or	
	Literal	B
	. . .	
	Or	
	Literal	Z
	Or	
	Literal	,
	Or	
	Literal	.
	Or	
	Literal	" "
	End of Table	
Numeric	Subroutine	3
	Literal	+
	Or	
	Literal	-
	Or	
	Literal	.
	Or	
	Literal	1
	. . .	
	Or	
Coordinate	Literal	%
	Repeat	Coord value
	Equation	Mode
	Subroutine	6
	End of Table	
Coord Value	Equation	Value
	Subroutine	4
	Repeat	Numeric
	Subroutine	5
	End of Table	
Mode	Subroutine	3
	Literal	A
	Or	
	Literal	D
	End of Table	

TABLE IV-2 (cont.)

Variable	Subroutine	3
	Literal	X
	Or	
	Literal	Y
	End of Table	
Data-Alpha	Literal	@
	Equation	Alpha
	Subroutine	7
	Repeat	Alphanumeric
	Subroutine	2
	End of Table	
Data-Numeric	Literal	#
	Equation	Alpha
	Subroutine	7
	Repeat	Numeric
	Subroutine	5
	End of Table	
Delimit	Literal	/
	End of Table	

Table IV-3

SUBR.1 Operation Codes

Operation Code 0

This operation allows the subroutine to initialize the machine data arrays and conversion subroutine parameters.

Operation Code 1

This code indicates that the input cursor points to the alphanumeric code for the data entry type. The conversion subroutine sets the type value in the machine buffers with values 0-3 assigned to point, line, area and grid data, respectively.

Operation Code 2

An alphanumeric string is terminated. Initially, VALUE points to itself which indicates to the subroutine that the first alphanumeric string is the name of the data entry. If this is the case, the subroutine returns to the driver. If VALUE does not point to itself, the alphanumeric string processed represents the descriptive value of an attribute in the attribute array. In this case, the address of the first character in the string which has been moved to the output buffer is placed in the machine location assigned to the attribute. For alphanumeric description valued attributes, the attribute location in the attribute array contains an address pointer to the description string.

Operation Code 3

The literal at the input cursor is moved to the output buffer and the output cursor is incremented.

Operation Code 4

The syntax driver has identified a coordinate variable at the input cursor. The pointer VALUE is set to the address of the coordinate buffer corresponding to the literal at the input cursor, "X" for x-coordinates and "Y" for y-coordinates. The pointer POINT is set to the current

Table IV-3 continued

value of the output cursor.

Operation Code 5

This operation code indicates that the character string moved to the output buffer is an ASCII coded decimal value. The subroutine converts this decimal integer into a binary integer and stores the binary number at the address indicated by the pointer VALUE.

Operation Code 6

The end of a coordinate data entry has been detected and the input cursor points to the coded value for the mode of the coordinate data point. If the coordinate mode is absolute, the values in the coordinate buffers are stored in the coordinate values array. If the coordinate mode is differential, the previous coordinates are added to the values in the coordinate buffers and the resultant coordinate values are stored in the coordinate stack.

Operation Code 7

This operation code indicated that an attribute variable has been indicated and the attribute code is at the input corner. The subroutine sets the VALUE pointer to the address of the specific attribute in the attribute array and sets the pointer POINT to the value of the output cursor.

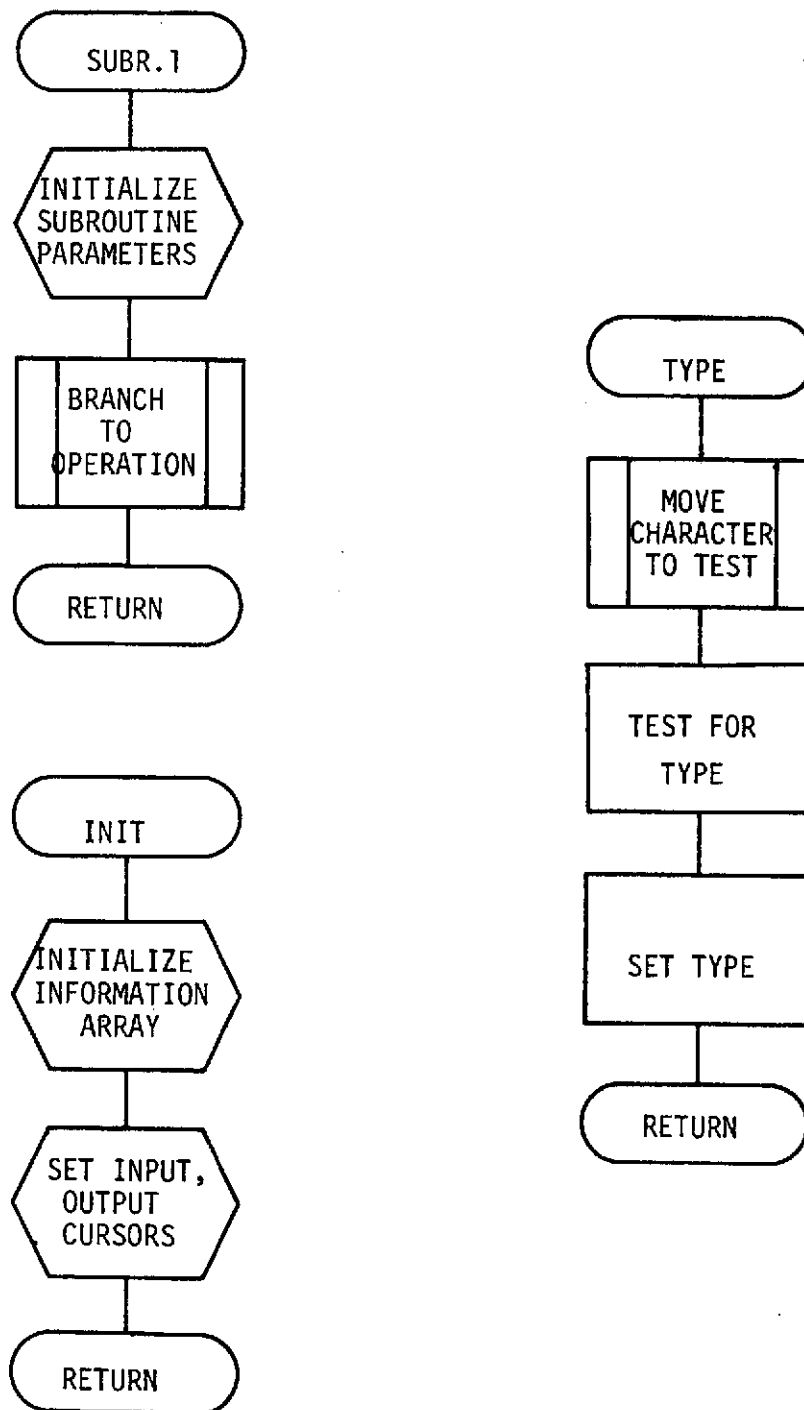


Figure IV-2. Logical Flow Diagram for SUBR.1

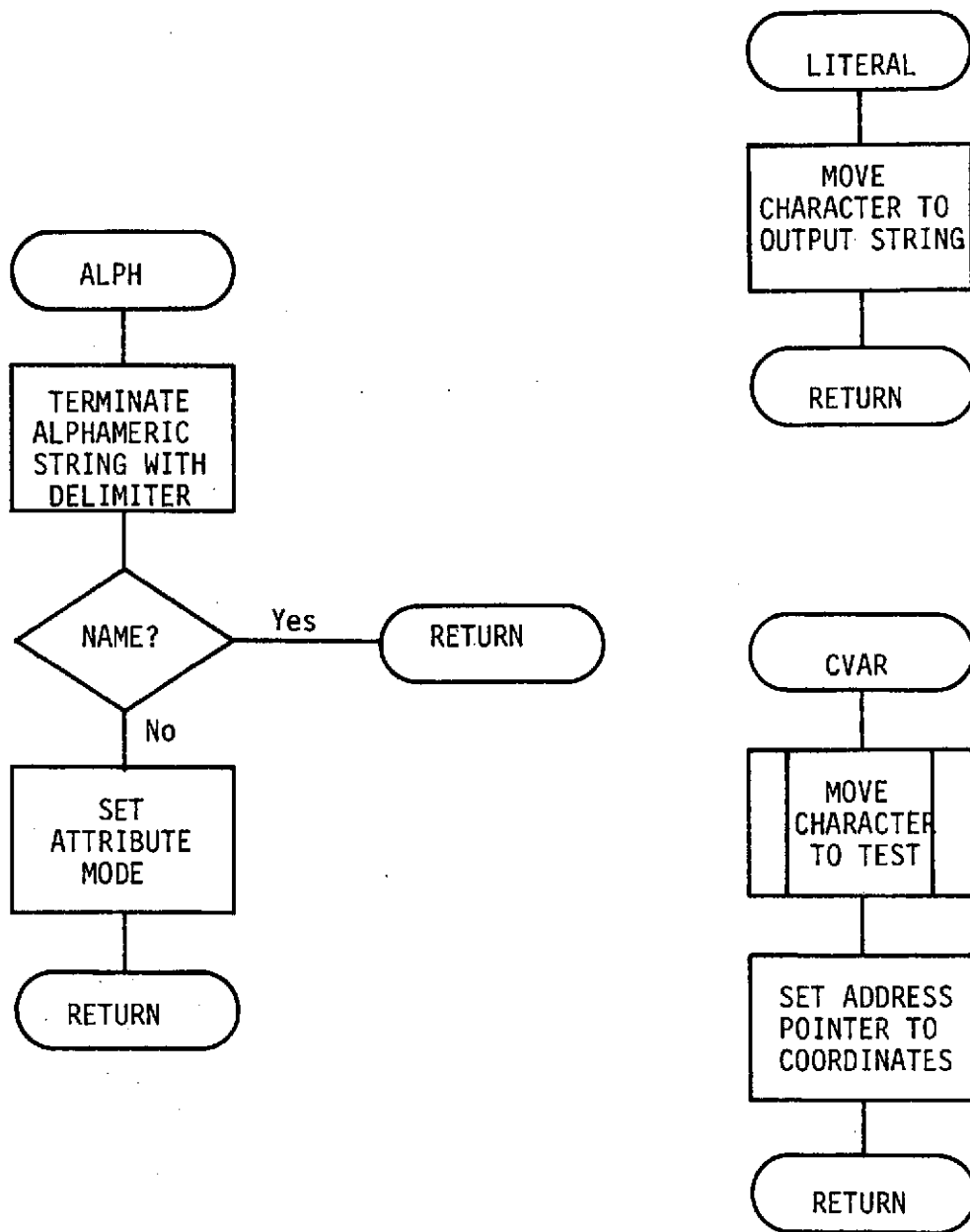


Figure IV-2. Logical Flow Diagram for SUBR.1 (cont.)

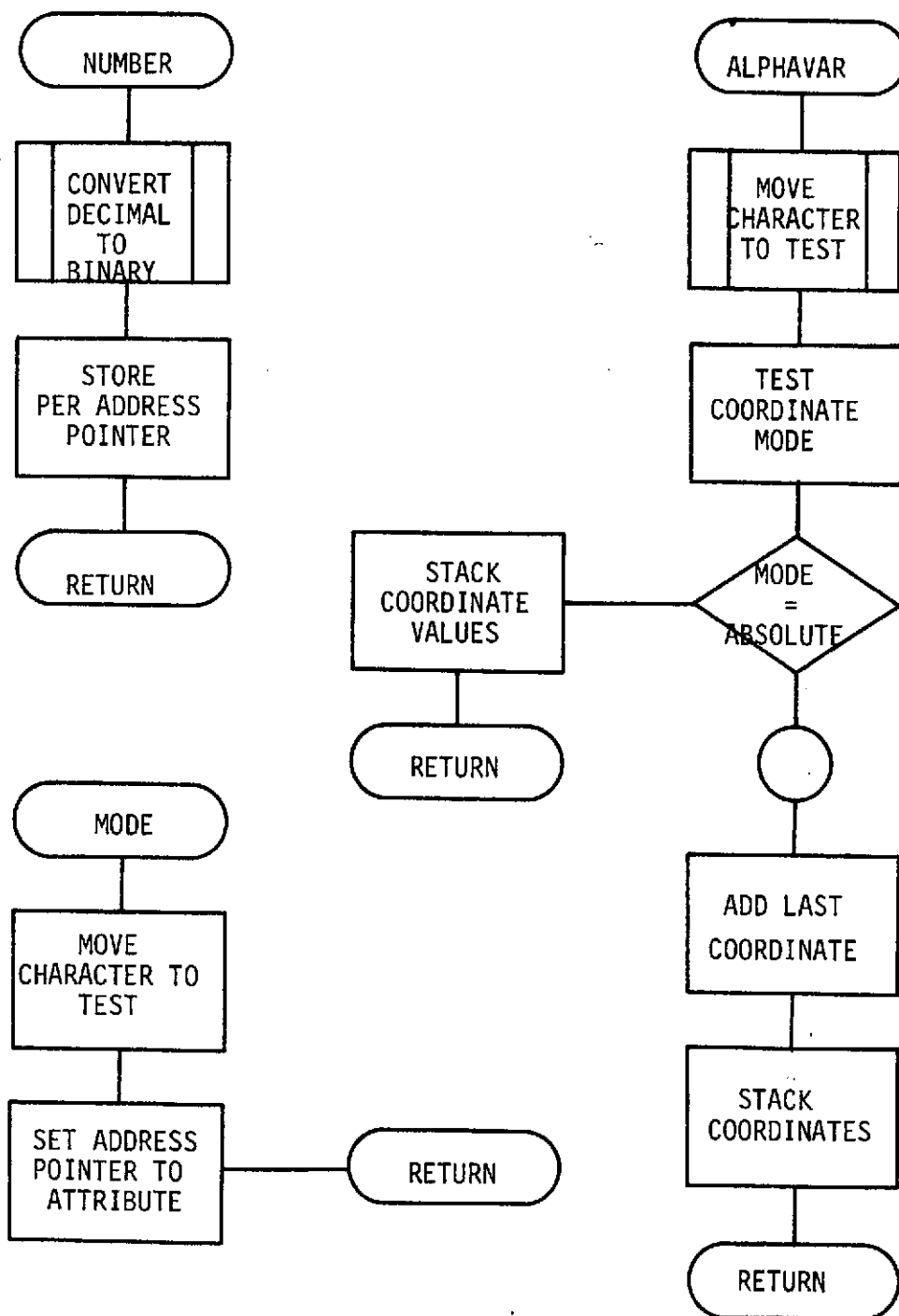


Figure IV-2. Logical Flow Diagram for SUBR.1 (cont.)

subroutine, input character text following the syntax definitions the equivalent machine variables are set according to the information input. Processing of the data in machine format may proceed from this point.

Data Transformation

Definition

A desirable feature incorporated into IRIMAS is the capability of combining weighted variables together in a functional form and the comparing of the variables and derived values. In order that these data transformations are programmable real-time, an interactive programming-like operator interface has been established. Program-like statements are input by a system operator and are interpreted by a syntax driver, syntax table and conversion subroutine. The output string from the conversion subroutine is interpreted to perform simple mathematical and logical functions on the machine format data.

Three specific sets of variables may be referenced: (1) a set of dummy variables; (2) the output variables which are kept separate from the input variables; and (3) the attribute values themselves. Each variable within the range of these arrays may be utilized by the interpretation scheme. The following types of data transformation, and logical operators are allowed:

Logical	Arithmetic
.EQ.	- subtraction
.NE.	+ addition
.GT.	* multiplication
.LT.	/ division
.GE.	** Exponentiation
.LE.	

Arithmetic subexpressions are also permissible in the data system as well as conditional statements like
 IF (A1 = 0.0) THEN (A2 = 3.0).

A limited number of variables are allowed for this interpretative phase. These variables are referenced by a prefix letter and a suffix number. Attribute variables are designated with the prefix "A" and twenty seven may be referenced (A0-A26). The first attribute variable (A0) designates the data entry type (point, area...etc.). A set of dummy variables are available for programming flexibility. These are designated with a prefix "D" and range from D0-D4. Similarly, output variables are designated with a previx "Ø" and range from Ø0-Ø4. Statements allowed are FORTRAN-like equations, however, no intrinsic functions (sine, cosine, etc.) are supported. Neither are any branch or input/output operations permitted. An example of a permissible set of statements is as follows:

```

Ø0 = 0
D0 = 3.0*A1 + 0.9 * A2 + 1*A3
D1 = A24 - A15/2
IF (D1.GT.D0) THEN (Ø0 = 255)
END.

```

The programming statements are defined by a syntax table and are converted into Polish algebra strings which are interpreted to effect the mathematical operations. Polish notation is described in [38]. Essentially, the technique presents numerical relationships as a string of numerical values and operators. It is useful in overcoming the confusion encountered with respect to mathematical operation priorities. In the example above, the statement:

$$D0 = 3.0 * A1 + 0.9 * A2 + 1 * A3$$

is found.

Confusion immediately arises as to how the sequence of operations should be applied. Does the statement imply operations grouped like:

$$D0 = (3.0 * A1) + (0.9 * A2) + (1 * A3),$$

or is the grouping,

$$D0 = 3.0 * (A1 + 0.9) * (A2 + 1) * A3,$$

appropriate?

Polish notation associates an operator with the two preceding data values. Computationally, an operation acts

upon the two preceeding data values in the string and the three are replaced in the string by the result of the operation. The computation begins with the first occurrence of an operator in the string and terminates when the supply of operators is exhausted. An algorithm for Polish notation interpretation generally keeps the data values in a stack until an operator is encountered. At this point, the last two data items in the operator stack are used together with the operator to compute the result which is returned to the stack. Polish representation of the above statement might be as follows,

$$D0 = 3.0 A1 * 0.9 A2 * + 1 A3 * +$$

and the computation of the sequence would be

$$D0 = (3.0 * A1) 0.9 A2 * + 1 A3 * +$$

$$D0 = (3.0 * A1) (0.9 * A2) + 1 A3 * +$$

$$D0 = ((3.0 * A1) + (0.9 * A2)) 1 A3 * +$$

$$D0 = ((3.0 * A1) + (0.9 * A2)) (1 * A3) +$$

$$D0 = (((3.0 + A1) + (0.9 * A2)) + (1 * A3))$$

Input to the analysis system for data transformation is accomplished through FORTRAN-like programming statements. These statements are restricted to arithmetic and logical operations on attribute, dummy and output variables. A syntax table (see Table IV-4) is used to define the statement input and when used with the syntax driver and a conversion subroutine with flow chart in Figure IV-3

TABLE IV-4
Data Transformation Syntax

Transform	Subroutine Equation End of Table	32 Statement
Statement	Equation Or Equation End of Table	Compound Simple
Compound	Literal Subroutine Equation Subroutine Literal Equation End of Table	IF 10 Expression 20 Then Simple
Simple	Equation Or Equation End of Table	Assign End
Assign	Subroutine Equation Literal Equation End of Table	11 Variable = Expression
End	Literal Subroutine End of Table	End 12
Expression	Equation Repeat End of Table	Priority 4 Priority 5
Priority 5	Literal Equation Subroutine Or Literal Equation Subroutine Or Literal	.EQ. Priority 4 13 .NE. Priority 4 14 .GT.

TABLE IV-4 (cont.)

	Equation	Priority 4
	Subroutine	15
	Or	
	Literal	.LT.
	Equation	Priority 4
	Subroutine	16
	Or	
	Literal	.GE.
	Equation	Priority 4
	Subroutine	17
	Or	
	Literal	.LE.
	Equation	Priority 4
	Subroutine	18
	End of Table	
Priority 4	Equation	Priority 3
	Repeat	Priority 4.1
	End of Table	
Priority 4.1	Literal	
	Equation	Priority 3
	Subroutine	21
	Or	
	Literal	+
	Equation	Priority 3
	Subroutine	22
	End of Table	
Priority 3	Equation	Priority 2
	Repeat	Priority 3.1
	End of Table	
Priority 3.1	Literal	*
	Equation	Priority 2
	Subroutine	23
	Or	
	Literal	1
	Equation	Priority 2
	Subroutine	24
	End of Table	
Priority 2	Equation	Priority 1
	Repeat	Priority 2.1
	End of Table	

TABLE IV-4 (cont.)

Priority 2.1	Literal Equation Subroutine	** Priority 1 25
Priority 1	Literal Subroutine Equation Or Literal Equation Or Equation End of Table	5 Operand + Operand Operand
Blank	Literal	"5"
Operand	Equation Or Equation Or Equation End of Table	Subexpression Variable Constant
Subexpression	Literal Equation Literal End of Table	6 Expression)
Constant	Equation Or Equation End of Table	Fraction Integer
Integer	Equation Repeat Subroutine End of Table	Number Number 0
Fraction	Equation Repeat Literal Subroutine Equation Repeat Subroutine Subroutine	Number Number . 0 Number Number 1 22

TABLE IV-4 (cont.)

End of Table		
Variable	Equation	Attribute
	Or	
	Equation	Dummy
	Or	
Attribute	Equation	Output
	End of Table	
	Literal	A
	Equation	Number
Dummy	Repeat	Number
	Subroutine	2
	End of Table	
	Literal	D
Output	Equation	Number
	Repeat	Number
	Subroutine	3
	End of Table	
Number	Literal	0
	Equation	Number
	Repeat	Number
	Subroutine	4
	End of Table	
	Subroutine	6
	Literal	0
	Or	
	Literal	1
	Or	
	Literal	9
	End of Table	

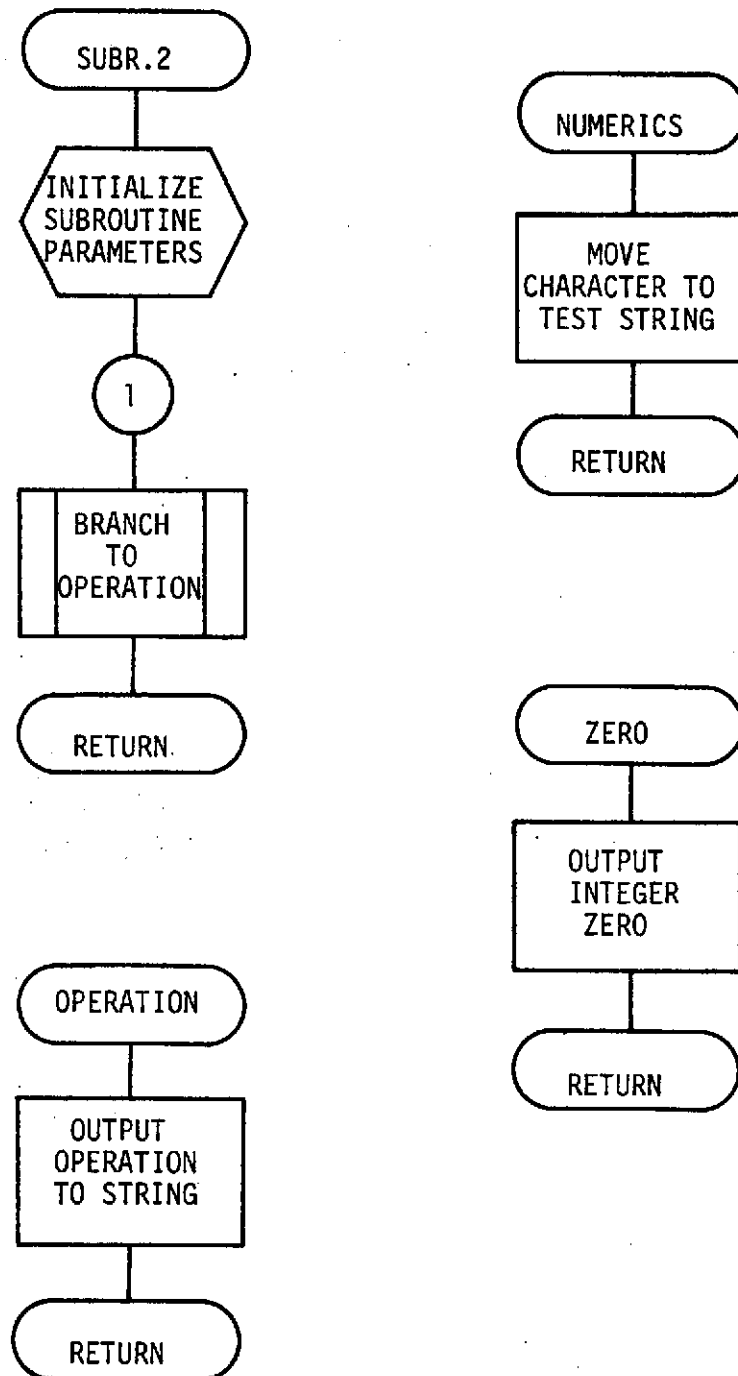


Figure IV-3. Logical Flow Diagram for SUBR.2

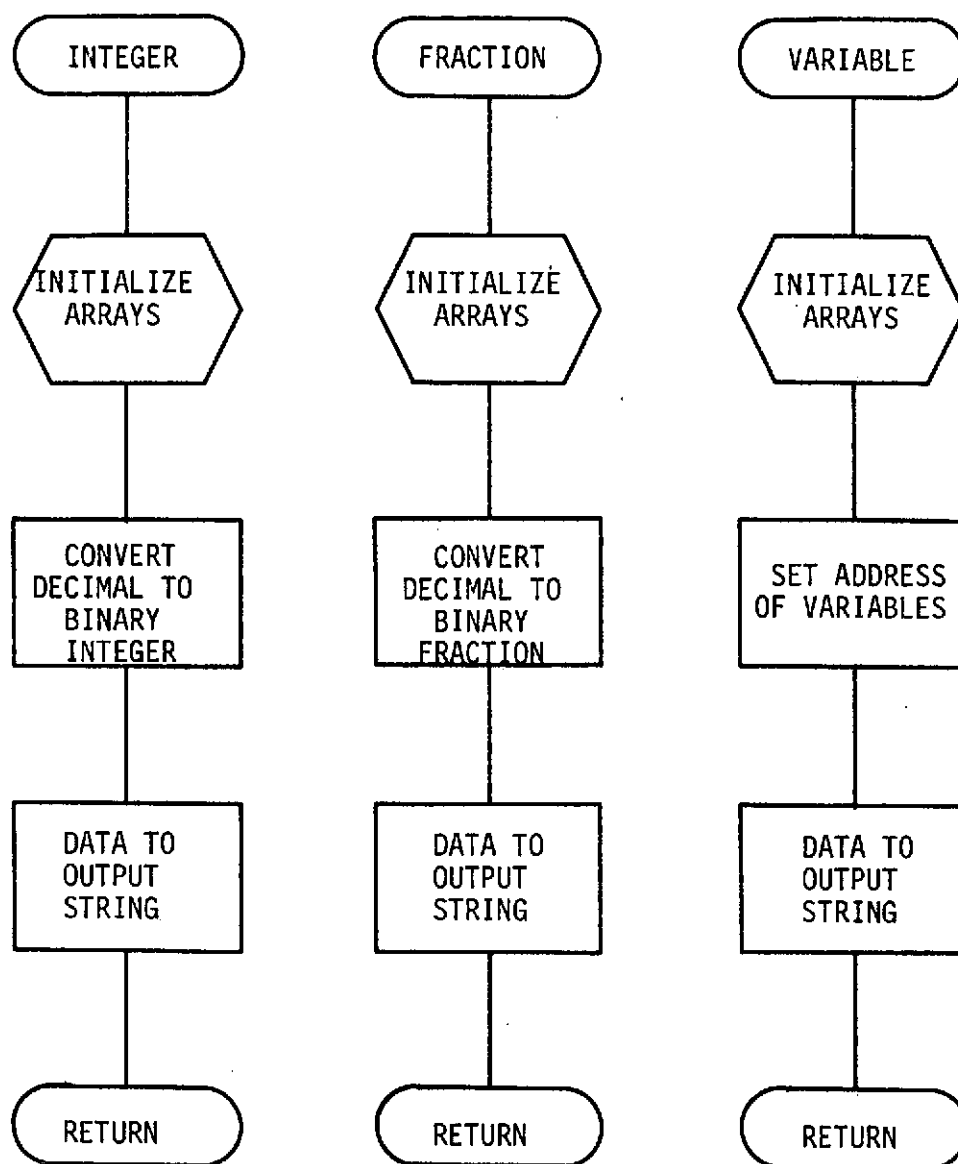


Figure IV-3. Logical Flow Diagram for SUBR.2 (cont.)

allows conversion of input statements to machine coded strings in Polish notation. An arithmetic interpreter subroutine is then used to effect the desired arithmetic operation by interpreting the Polish string prepared by the syntax procedures.

Conversion Subroutine

The conversion subroutine SUBR.2 is utilized with the syntax driver to convert the interpreted syntax into a Polish form string for subsequent computation. The syntax table provides the operation codes for the conversion subroutine shown in Table IV-5 and the functions performed by the subroutine are listed in Table IV-6. Table IV-7 summarizes the 8-bit operation codes placed in the output string as a result of the syntax interpretation.

Transformation Interpreter

The transformation interpreter TRANS interprets the Polish format string provided by the IRIMAS syntax transformation conversion subroutine, and executes the indicated operations. A byte string of operation codes, identifiers, integer and fractional binary values and variable addresses constitute the input string processed by the interpreter. Pseudo-floating point operations are performed where the binary point is shifted to combine

Table IV-5

Syntax Conversion Subroutine SUBR.2 Operation Codes

0	Integer number	13	.EQ.
1	Fraction	14	.NE.
2	Attribute variable	15	.GT.
3	Dummy variable	16	.LT.
4	Output variable	17	.GE.
5	Constant zero	18	.LE.
6	Numerics	19	-----
7	-----	20	STOP
8	-----	21	-
9	-----	22	+
10	IF	23	*
11	ASSIGN	24	/
12	END	25	**

Table IV-6

Subroutine SUBR.2 Interpretation Functions

Operation Code 1

The ASCII coded decimal fraction transferred to the LIST buffer area is converted to a binary fraction. The binary fraction, preceded by a "0" for identification, is placed in the output string.

Operation Code 2, 3, 4

These opcodes indicate that an attribute, dummy, or output variable has been referenced. These variables are maintained in arrays and the ASCII coded decimal number in the LIST buffer area indicates which member of the array is being referenced. The number is converted to a binary integer and the appropriate variable address is calculated. The variable address, prefixed with a value "02" to specify a variable address, is inserted in the output string.

Operation Code 5

A value of zero, prefixed with an integer identification "00", is placed in the output string.

Operation Code 6

The character referenced by the input cursor moved to the character string at position LIST.

Operation Code 7-25

Mathematical, logical, and control functions are represented by operation codes 7-25. These same functions are required by the computation software, therefore, the conversion subroutine simply places the codes in the output string.

Table IV-7
Polish String Operation Codes

Code	Function
0	A 16-bit binary integer follows
1	A 16-bit binary fraction follows
2	A 16-bit address follows
7-9	No operation specified
10	Conditional expression (IF)
11	Assign
12	End
13	Logical equal (.EQ.)
14	Logical not-equal (.NE.)
15	Greater than
16	Less than
17	Greater than or equal
18	Less than or equal
19	No operation specified
20	Stop
21	Subtract
22	Add
23	Multiply
24	Divide
25	Exponentiate

integer and fraction values which can be handled by integer arithmetic operations. The specific operation codes and data identifiers are discussed in previous sections and are summarized in Tables IV-5 (page 104) and IV-7 (page 106).

Coordinate Transformation

Transformation Definitions

Three fundamental coordinate transformations are implemented by IRIMAS software. These include coordinate translation where a coordinate reference system is displaced along the x and y axis; coordinate rotation where the coordinate reference system is turned through a specified angle; and scaling whereby the coordinate reference system is expanded or contracted along the x and y axis. In Figures IV-4a through IV-4d are shown, by example, the results of these three coordinate transformations.

Coordinate translation is described by the following relationships:

$$x' = x + T_x \text{ and,}$$

$$y' = y + T_y,$$

where x' and y' are the new coordinates and x and y are the coordinates which are being transformed. In the translational mode the value of the coordinate x is displaced by the amount of T_x and the value of the

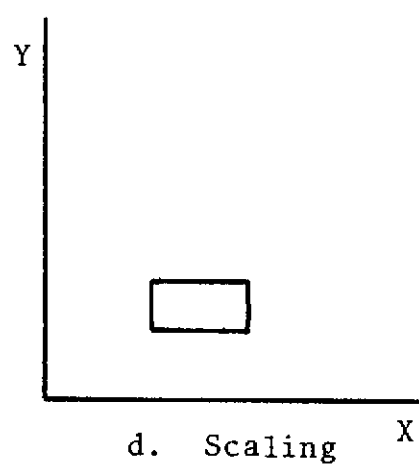
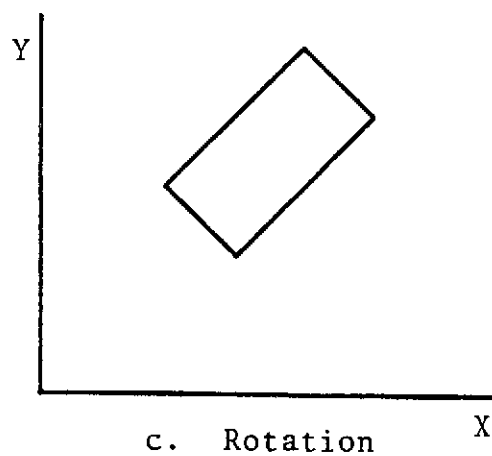
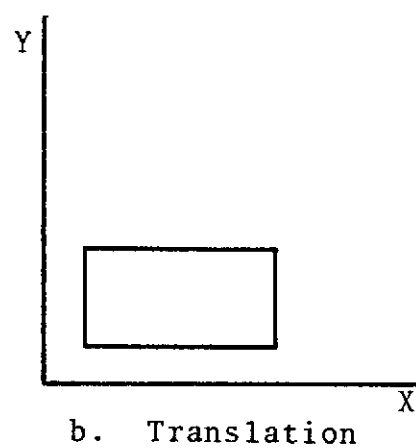
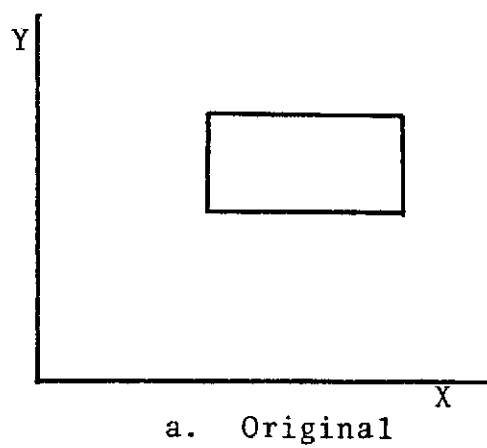


Figure IV-4. Coordinate Transformation Example

coordinate y is translated by the amount of T_y . Coordinate transformation may be conveniently handled as matrix operations or vectors. For the translation case, the vector-matrix product becomes

$$\begin{Bmatrix} x' \\ y' \\ 1 \end{Bmatrix} = \begin{Bmatrix} 1 & 0 & T_x \\ 0 & 1 & T_y \\ 0 & 0 & 1 \end{Bmatrix} \begin{Bmatrix} x \\ y \\ 1 \end{Bmatrix},$$

where the dummy value "1" has been introduced for operational convenience. Another convenience provided by the matrix notation arises from the understanding that a transformation is a single entity or operation, and that transformations can be combined or concatenated to yield a single transformation. Utilization of the matrix format allows ready reduction of the transformation string by application of the matrices on each other until a single matrix results. The transformation operator for translation then is

$$T_T = \begin{Bmatrix} 1 & 0 & T_x \\ 0 & 1 & T_y \\ 0 & 0 & 1 \end{Bmatrix}$$

The second coordinate transformation to be considered is that of rotation. Coordinate rotation is described by the relationship:

$$x' = x \cos \theta + y \sin \theta$$

$$y' = -x \sin \theta + y \cos \theta$$

where θ is the angle through which the coordinate system is to be rotated. In matrix notation, the rotation transformation becomes:

$$T_R = \begin{Bmatrix} \cos \theta & \sin \theta & 0 \\ -\sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{Bmatrix}$$

The final coordinate transformation considered is that of scaling. Coordinate system scaling is described by the relationships:

$$\begin{aligned} x' &= x S_x \\ y' &= y S_y \end{aligned}$$

where S_x and S_y are the scale factors applied to the x-axis and y-axis respectively. In matrix notation, the transformation is described by:

$$T_S = \begin{Bmatrix} S_x & 0 & 0 \\ 0 & S_y & 0 \\ 0 & 0 & 1 \end{Bmatrix}$$

Once concatenated matrices have been reduced to a single transform operation, the dummy variable need not be calculated in the final result. Thus, the transformation operation becomes

$$\begin{Bmatrix} x' \\ y' \end{Bmatrix} = \begin{Bmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \end{Bmatrix} \begin{Bmatrix} x \\ y \\ 1 \end{Bmatrix}$$

Syntax and Interpretative Subroutine

Operator specification of coordinate transformations follow syntactic definitions in the same manner as data transformation. Data entries follow the format $TR(T_x, T_y)$ for coordinate translation, $ROT(\theta)$ for rotation and $SCL(S_x, S_y)$ for scaling. The additional directive `END` terminates the operator entity. Examples of allowable operator entries are:

TR (10,0)	translate
ROT (25)	rotate 25 degrees
TR (-10,0)	translate
SCL (10.0, 0.10)	scale
END	terminate operation

The formal syntax for coordinate transformation is defined in Table IV-8.

An interpretative subroutine, SUBR.3, is implemented to convert interpreted syntax into a coded string. Its flow diagram is in Figure IV-9. This coded string is used by the TRANS subroutine (see Figure IV-6) to execute the coordinate transformations defined by the system operator. The matrix operations previously described are implemented in data strings within the SUBR.3 subroutine. As the coordinate transformations defined by the operator are interpreted, the appropriate transformation parameters are inserted into the coded strings. Subroutine operations

TABLE IV-8
Syntax for Coordinate Transformation

Transform	Subroutine	0
	Equation	Operation
	Or	
	Equation	End
	End of Table	
Operation	Equation	Translate
	Subroutine	7
	Or	
	Equation	Rotate
	Subroutine	7
	Or	
	Equation	Scale
	Subroutine	7
	End of Table	
Translate	Literal	TRL (
	Equation	Integer
	Subroutine	1
	Literal	,
	Equation	Integer
	Subroutine	2
	Literal)
	End of Table	
Rotate	Literal	ROT (
	Equation	Integer
	Subroutine	3
	Literal)
	End of Table	
Scale	Literal	SCL (
	Equation	Integer
	Literal	.
	Equation	Fraction
	Subroutine	4
	Literal	,
	Equation	Integer
	Literal	.
	Equation	Fraction
	Subroutine	5
	Literal)
	End of Table	

TABLE VI-8 (cont.)

Integer	Equation	Number
	Repeat	Number
	Subroutine	9
	End of Table	
Fraction	Equation	Number
	Repeat	Number
	Subroutine	10
	End of Table	
Number	Subroutine	6
	Literal	0
	Or	
	Literal	1
	Or	
	Literal	9
	Or	
	Literal	-
	Or	
	Literal	+
	End of Table	
End	Literal	End
	Subroutine	8
	End of Table	

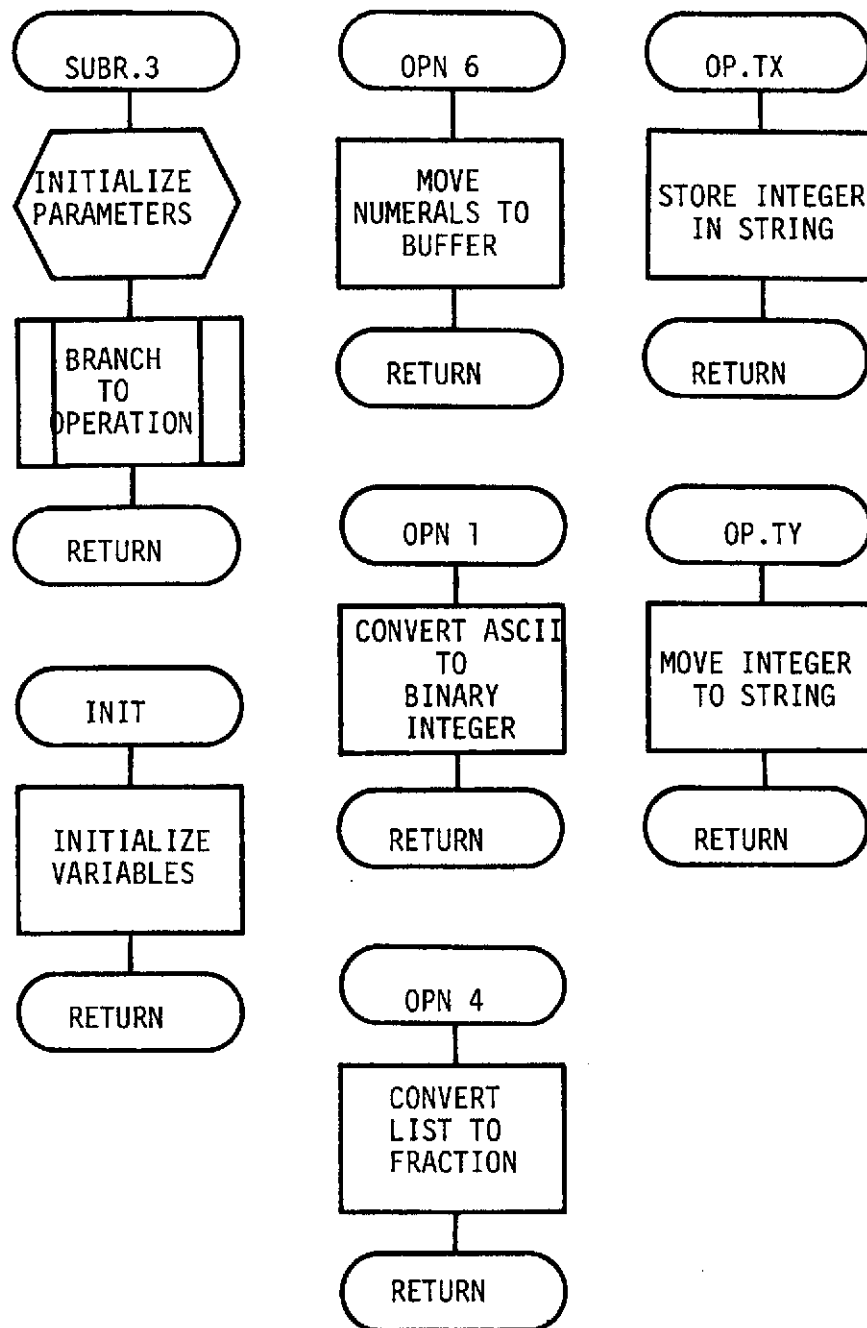


Figure IV-5. Logical Flow Diagram for SUBR.3

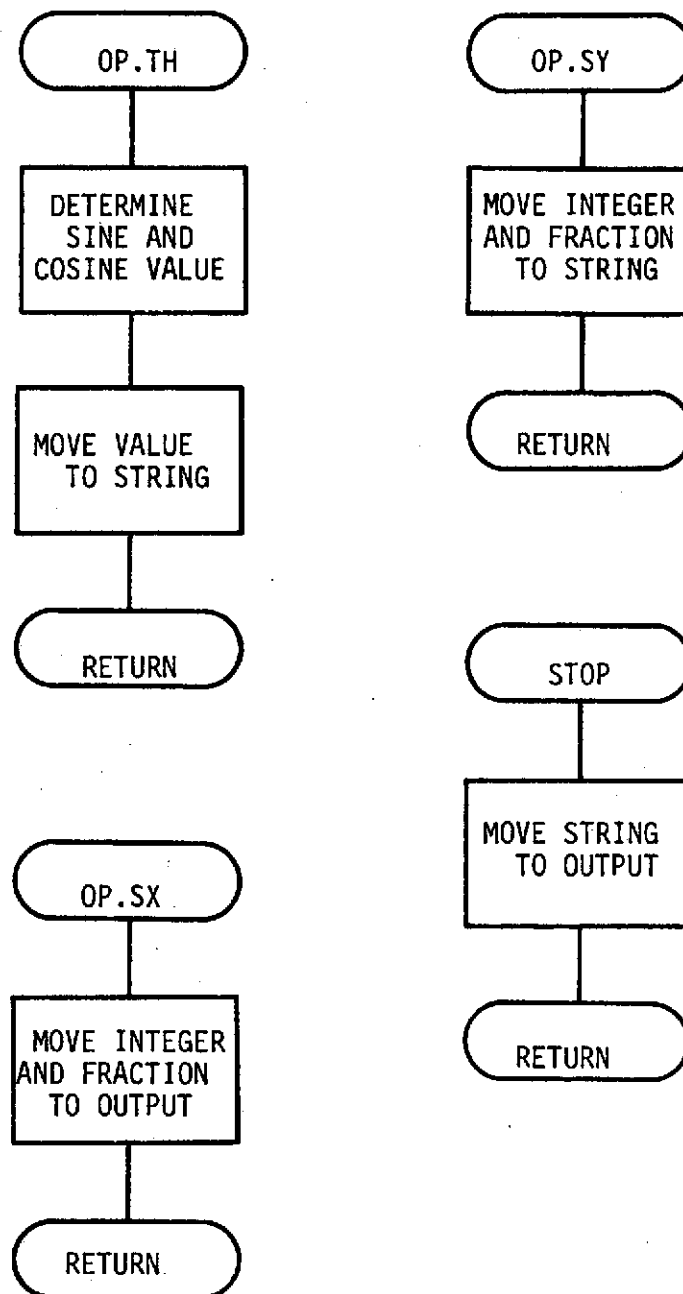


Figure IV-5. Logical Flow Diagram for SUBR.3 (cont.)

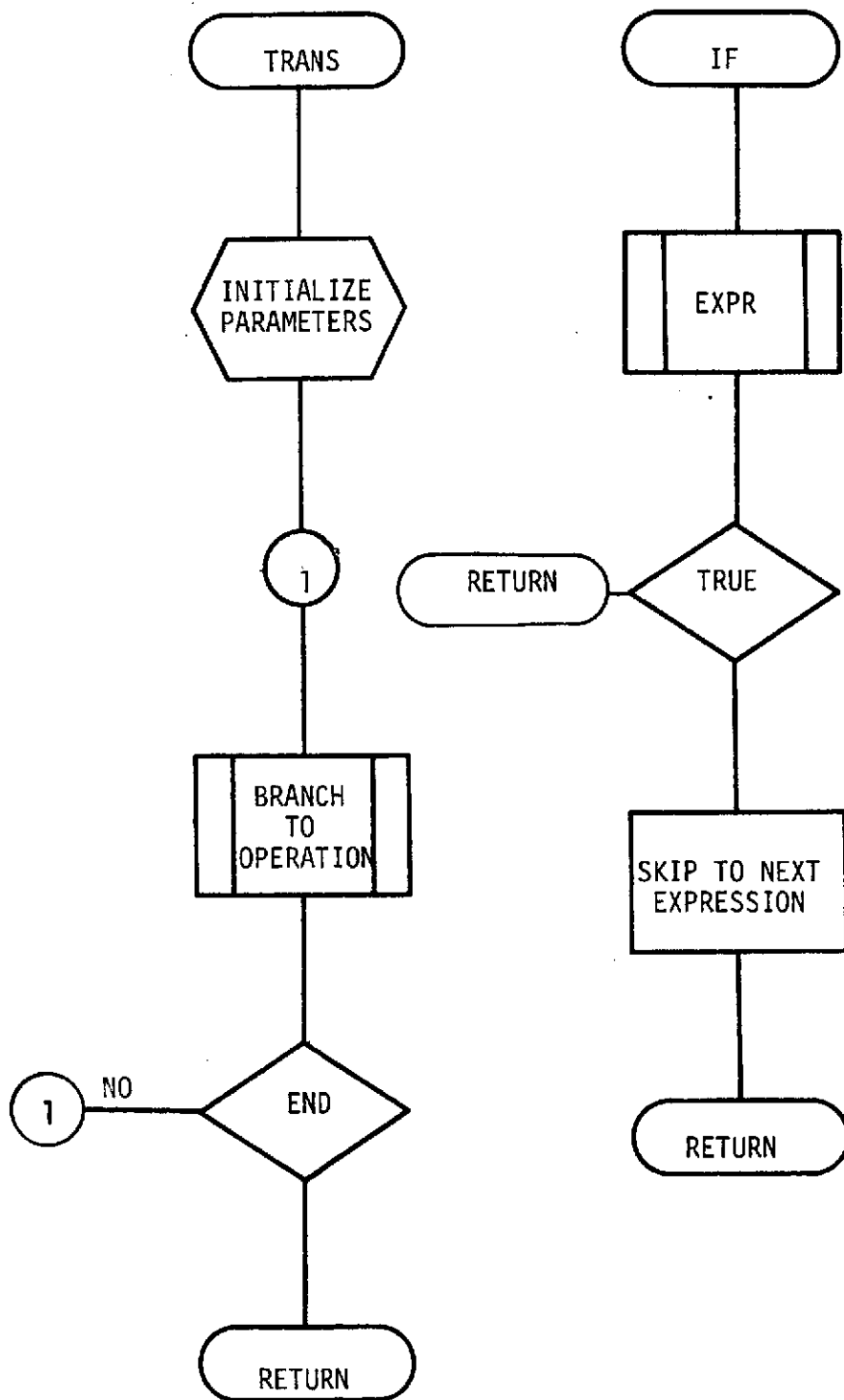


Figure IV-6. Logical Flow Diagram for TRANS

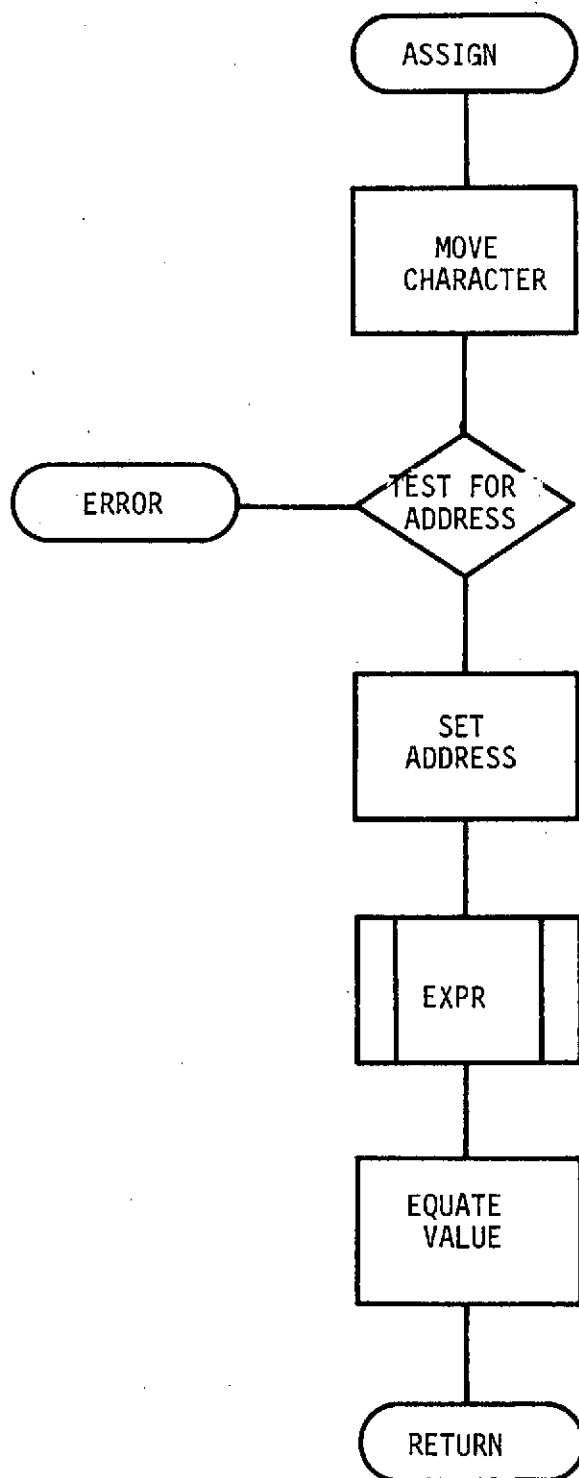


Figure IV-6. Logical Flow Diagram for TRANS (cont.)

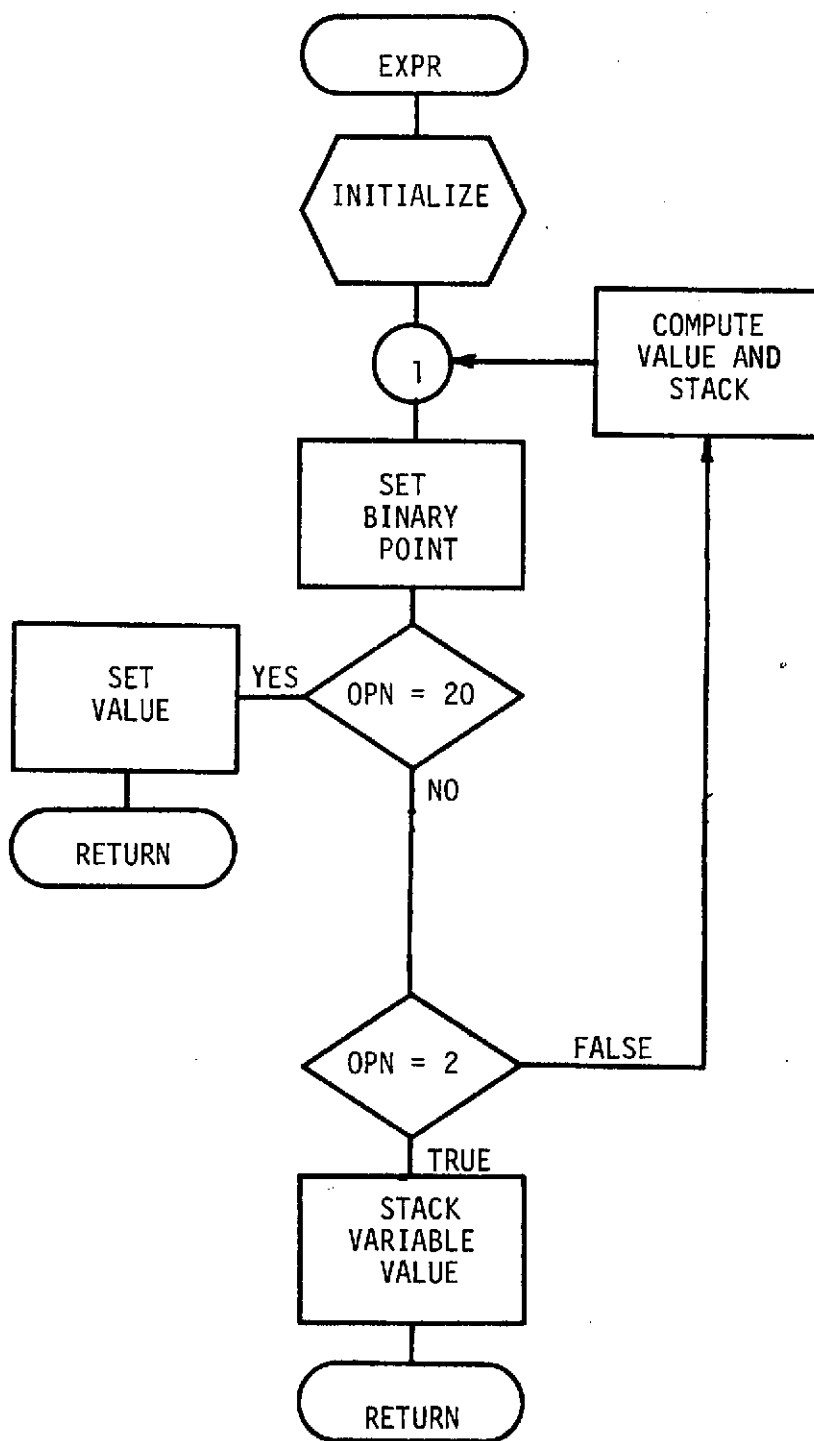


Figure IV-6. Logical Flow Diagram for TRANS (cont.)

codes are tabulated in Table IV-9.

Data Display

Man is accustomed to thinking in pictures and thus images provide him with a very personal view of his world. The use of thematic cartography is invaluable to the regional, natural resource, or land-use planner dealing with the spatial extent of land-use. The capability of image presentation is vital to the success of a resource information system.

Three graphical display devices are commonly available for image presentation, the line printer, the digital plotter and a cathode ray tube (CRT) display. The line printer and the CRT display, for the most part, print or plot data on an x-y coordinate grid in a sequential manner rather than by graphical interpretation of coordinate functions. Simply, the grid or raster data format is most compatible with these devices. In view of this requirement, IRIMAS software also converts the machine representation of the boundary format into the grid format.

Several functions are important to the format conversion of the boundary format data. Algorithms are available in IRIMAS for the tracing of boundary lines into an image

Table IV-9

Syntax Conversion Subroutine SUBR.3 Operations

Operation Code 0

The interpreting subroutine is initialized.

Operation Code 1

The value for an x-coordinate translation, T_x , is interpreted.

Operation Code 2

The value for a y-coordinate translation, T_y , is interpreted.

Operation Code 3

The value for the angle of a coordinate rotation in degrees, θ , is interpreted.

Operation Code 4

The value for an x-coordinate scale change is interpreted.

Operation Code 5

The value for a y-coordinate scale change is interpreted.

Operation Code 6

The input cursor points to a character in an ASCII coded decimal number which is moved into a buffer area for conversion.

Operation Code 7

The coordinate transformation specified has been completely translated.

Operation Code 8

The end of the definition phase for coordinate transformation is indicated.

Operation Code 9

The ASCII coded character string of numerals in the buffer storage area is converted to a binary integer.

Operation Code 10

The ASCII coded character string of numerals in the buffer storage area is converted to a binary fraction.

buffer area, as well as for shading the area within the boundary. These algorithms convert the lineal and area boundary format data into grid data where coordinates are associated with memory locations. The point-to-grid format conversion is direct. An additional algorithm is provided to clip the boundaries of the information for display to fit the area defined by the display extent.

Clipping

A clipping algorithm is implemented in IRIMAS which rejects those portions of the data boundaries which lie off of the window area defined by the display. The algorithm additionally computes the endpoints of the boundary lines at the intersection with the display boundaries effectively redefining the data boundary to lie entirely within the display window. The implemented clipping algorithm, attributed to Cohan and Sutherland by Newman et. al. [35], not only rapidly finds the endpoints off the window but is also able to quickly reject any line which lies entirely off the display area.

The algorithm is in two parts; the first part applies a rejection test. This test is implemented by extending the edges of the window so that the space occupied by the unclipped picture is divided into nine regions. Each of these regions has a four bit code as shown in Figure IV-7.

1001	1000	1010
0001	0000	0010
0101	0100	0110

Figure IV-7. Logical Codes for Clipping Algorithm

The endpoints of the line are assigned codes according to the region in which they fall. The four bits in the code mean the following if set:

- first bit: point is above top edge of display
- second bit: point is below bottom edge
- third bit: point is to the right
- fourth bit: point is to the left.

If the four bit codes for both endpoints are zero, the line lies in the window, and if the intersection of the two codes is not zero, the line must lie entirely off the display.

If the line cannot be eliminated by either of these tests, then the second portion of the algorithm is applied. A line which lies partially out of the display area can be divided into two parts, one of which must lie entirely outside the display area. A simple method of subdivision is to find the line's point of intersection with one of the window edges and to throw away that part which can be rejected. For example, the line AB in Figure IV-8 could be subdivided at C, and applying the rejection tests the portion AC can be discarded. The line BC cannot be simply categorized and therefore must be subdivided again at D. The result BD is found to lie entirely inside the window area. In applying the rejection test repeatedly, the order of the rejections does not affect the final result.

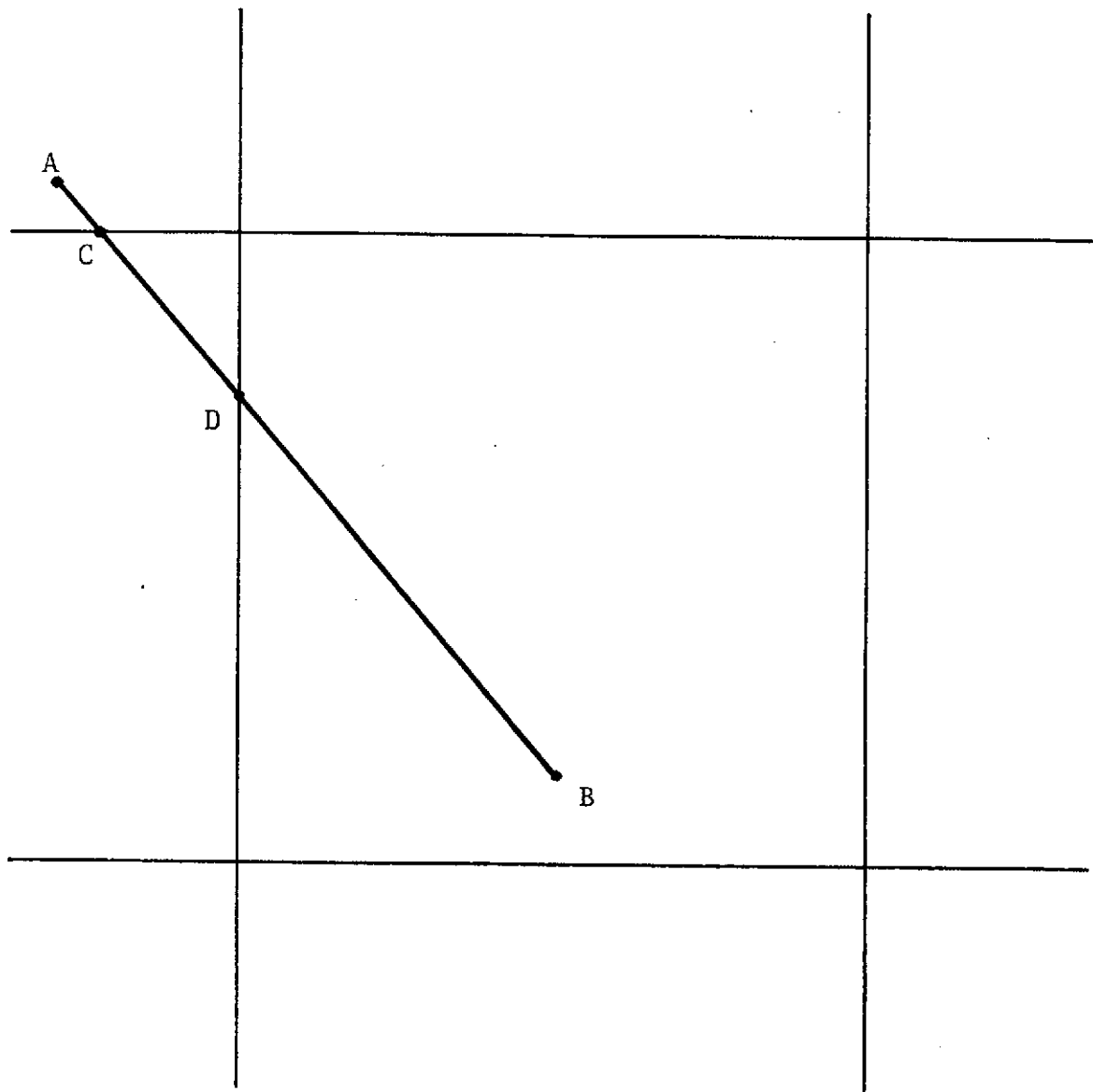


Figure IV-8. Example of Line Rejection and Clipping

Line generation

Information system data entries handled by IRIMAS are represented as piecewise linear boundaries specified by line segment endpoints. An algorithm is implemented for expansion of the endpoint coordinates into the set of coordinates which constitute the line. The algorithm is based on parametric representations of the x and y coordinates of a line. The set of parametric equations are expanded from endpoint to endpoint as the coordinates of the intermediate points on the line are calculated.

A line with endpoints (x_1, y_1) and (x_2, y_2) can be represented by a set of parametric equations:

$$x = x_1 + (x_2 - x_1)t$$

$$y = y_1 + (y_2 - y_1)t$$

where $0 \leq t \leq 1$.

At the starting point of the line $(x = x_1, y = y_1)$ t equals 0 and at the termination of the line segment $(x = x_2, y = y_2)$ t equals 1. The line representation is constrained by the grid coordinate representation to have coordinate values at discrete integer valued points. Thus, the line becomes a set of integer points. This suggests that the parameter " t " can assume only discrete values corresponding to the discrete coordinates.

In tracing the line, the algorithm generates two values of " t " for each additional coordinate point. The

value $t = t_1$ is computed by incrementing the x-value of the previous coordinate by Δx and calculating

$$t_1 = \frac{(x - x_1) + \Delta x}{(x_2 - x_1)} .$$

This computes a value of the parameter t keyed to the movement along the x-axis. A value for $t = t_2$ is computed by incrementing the y-value by Δy

$$t_2 = \frac{(y - y_1) + \Delta y}{(y_2 - y_1)} .$$

This value for t is keyed to movement along the y-axis. Since t is a monotonically increasing value from the initial coordinate (x_1, y_1) the coordinate point of the line closest to the preceding point has the smaller t value. Thus, the next point in the coordinate sequence tracing the line is generated by the smaller of $[t_1, t_2]$.

Since the coordinate sets (x_1, y_1) and (x_2, y_2) are arbitrary, movement from x_1 to x_2 , and y_1 to y_2 can require either positive or negative increments depending upon the relative relationships of the coordinates. Thus, the value of Δx and Δy used to increment the line coordinates can either be positive or negative depending upon the particular direction of the line. Appropriately chosen, the parameter t remains positive valued between 0 and 1.

Area Generation

The display of area information requires that the coordinate values of points within data boundaries be identified and associated with the information values of the area. In IRIMAS, an algorithm is provided to determine those points internal to a boundary and to compute the intersection of these points with the display window.

This "shading" algorithm computes the intersection of the line $y = y_i$, with the area boundary segments for values of y_i within the upper and lower y-limits of the line endpoints. These intersections are compared with each line segment's endpoints to determine if the intersection is on the boundary of the area. These x-coordinate values are sorted in ascending order and are compared with the x-axis limits of the display area. Those points without the display area are replaced by the display limits. The resultant set of x-axis values, taken in pairs, provide upper and lower limits of the points along the line $y = y_i$ which are both in the bounded area as well as in the display window. In Figure IV-9 is shown a graphic example of this shading performed by the algorithm. This algorithm provides a method for identifying the points encompassed by the data entry area which lie with the display area in order that area information can be transformed into the grid format.

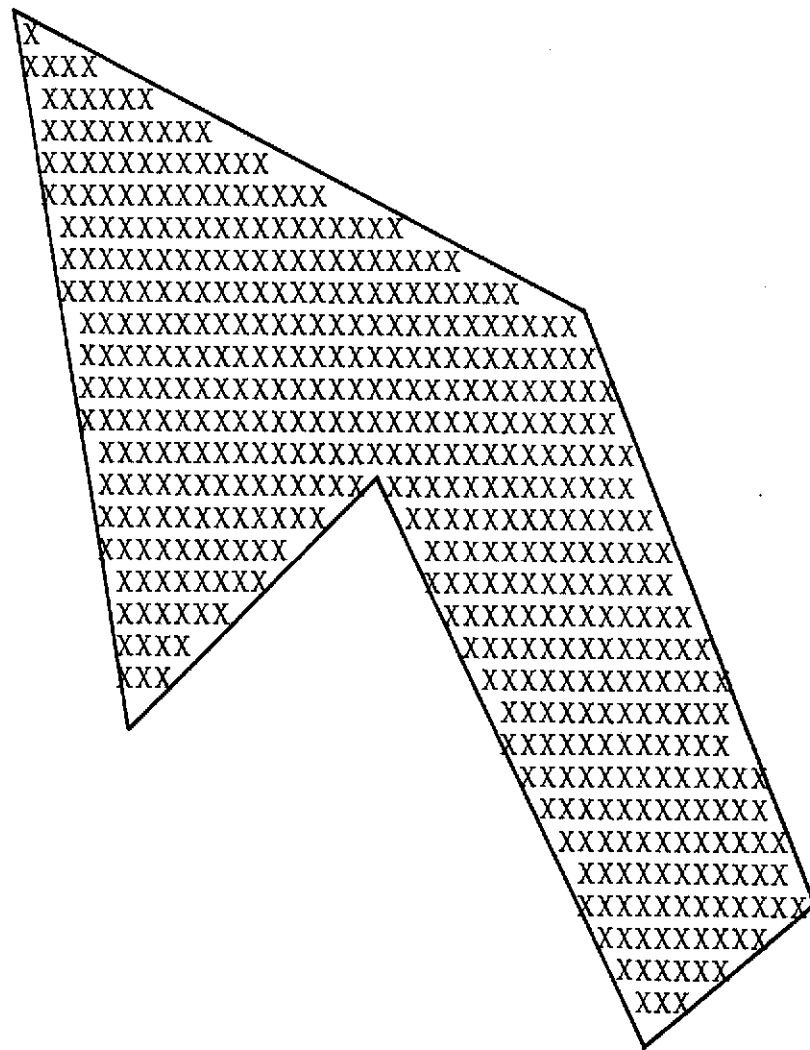


Figure IV-9. Example of Algorithmic Shading

Expansion Subroutine

An IRIMAS expansion subroutine EXPAND is implemented to convert the machine format data into the grid format for display and image-function operations. This subroutine implements the described algorithms for clipping, line generation and area generation and logically intersects the generated coordinate values with specified coordinate bounds. If a calculated data point lies within the specified coordinate frame, the subroutine execution is interrupted. At this point, the data values may be inserted into a coordinate indexed array thus completing the expansion. In Figure IV-10 is shown a flow graph for EXPAND.

Optionally, the subroutine interruption may be utilized as an indication of congruence between the machine format data set and an additional data set used to define the data frame input to EXPAND. For example, rectangular classification training sets defined as data frame boundaries cause an interrupt to occur when the input data corresponds to the training set areas. Under this condition, data statistics could be computed for utilization in classification algorithms. Alternatively, coordinate indexed locations in an array could be tested against a preset logical condition to control further processing. The subroutine EXPAND provides for convenient conversion

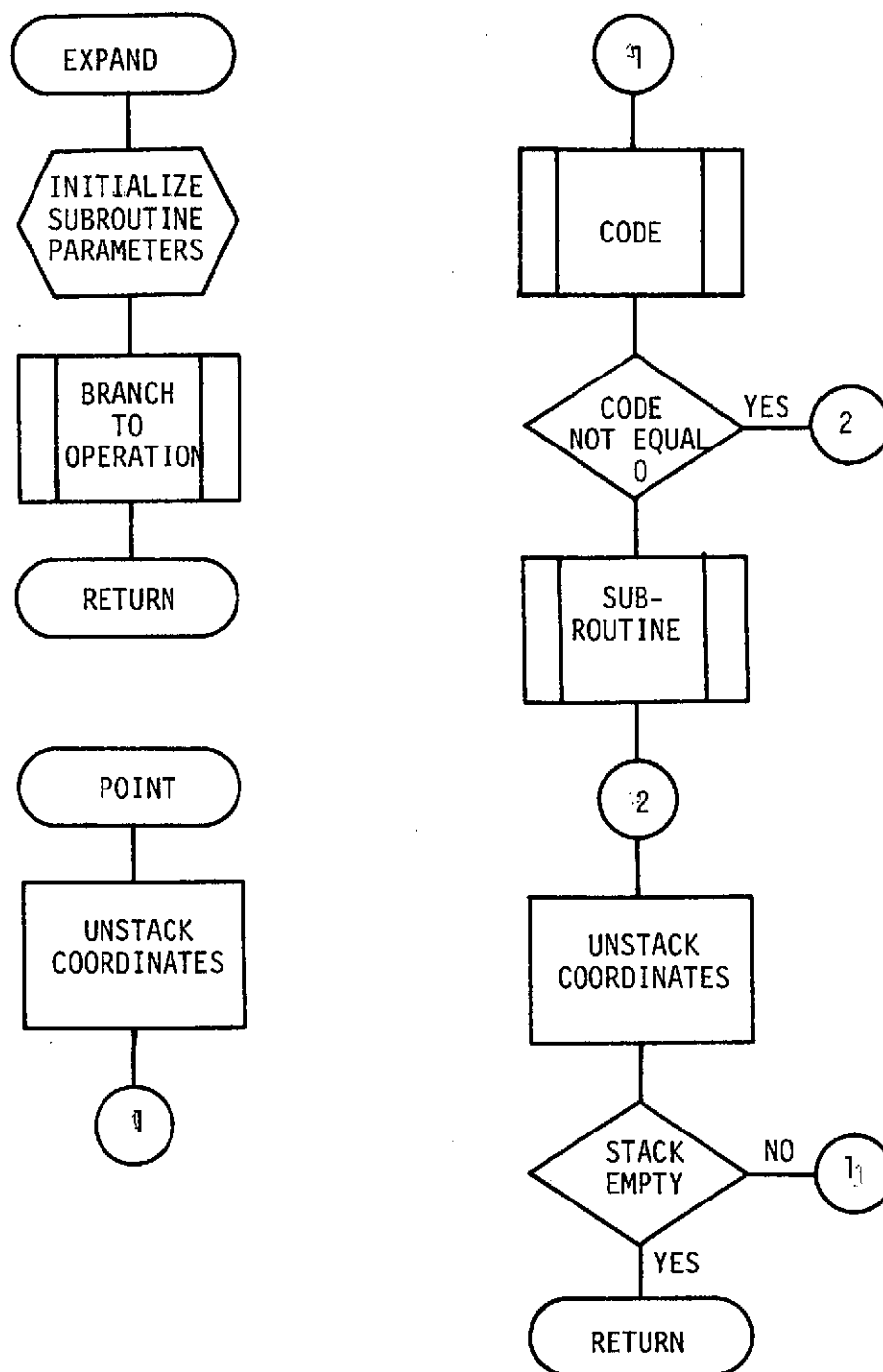


Figure IV-10. Logical Flow Diagram for EXPAND

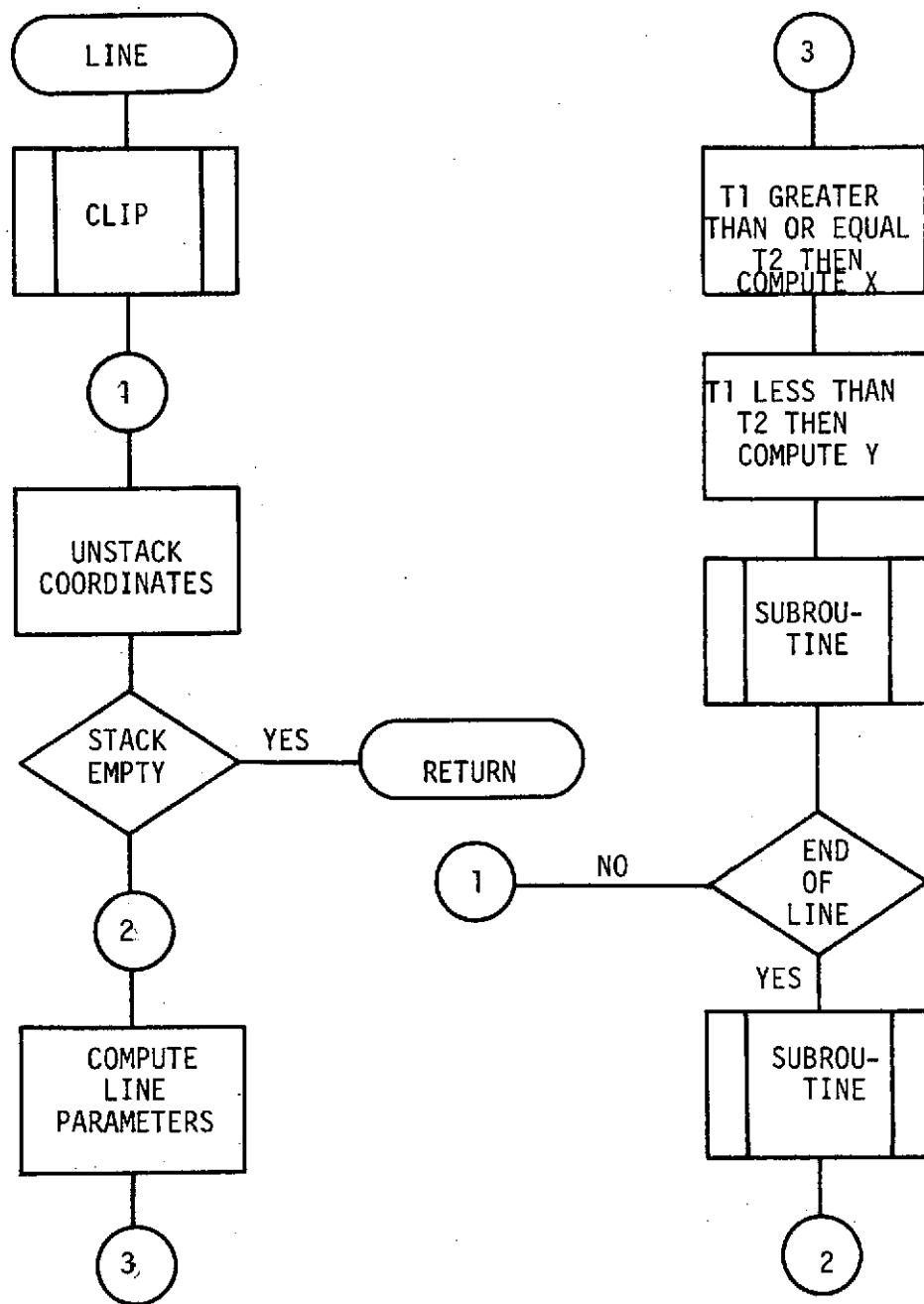


Figure IV-10. Logical Flow Diagram for EXPAND (cont.)

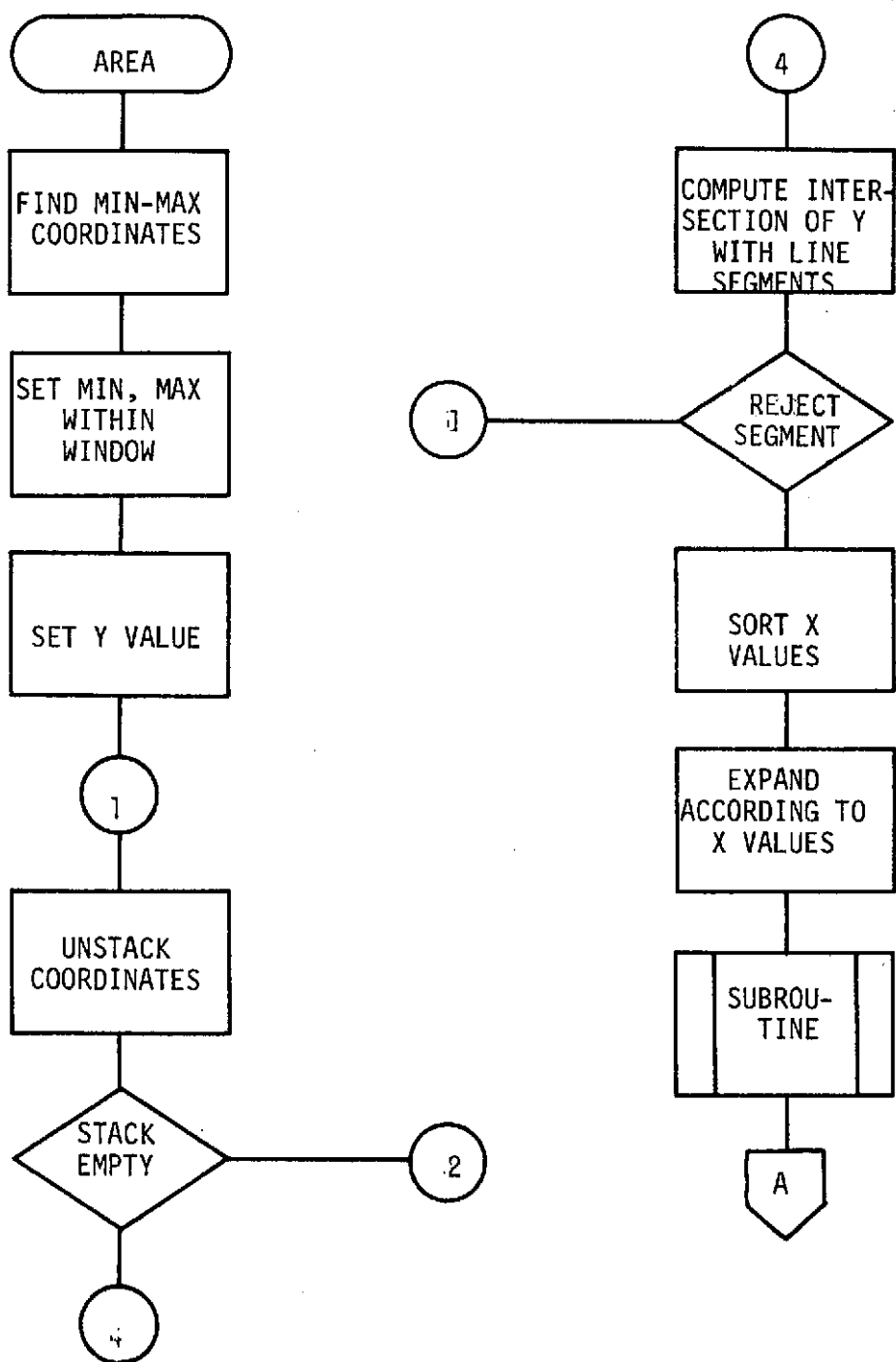


Figure IV-10. Logical Flow Diagram for EXPAND (cont.)

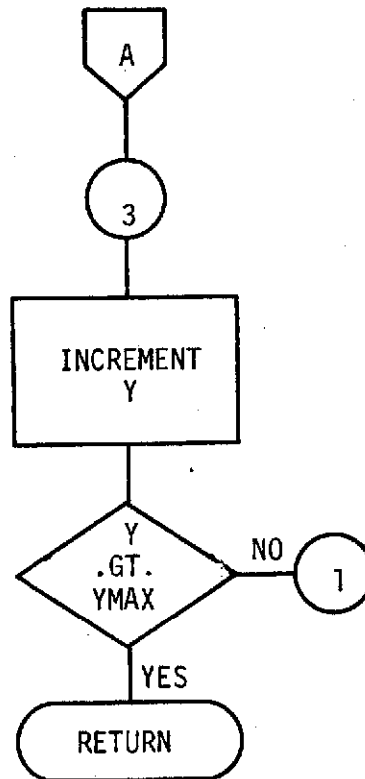


Figure IV-10. Logical Flow Diagram for EXPAND (cont.)

of the machine representation of boundary data to coordinate indexed arrays for further image processing or display.

CHAPTER V

IMAGE PROCESSING

The current availability of timely synoptic imagery from large portions of the productive land and ocean areas of the world offers an unprecedented source of information to the resource manager. Properly interpreted, this information can provide new insight into resource availability and productivity and especially the environmental impact of human interaction with natural ecosystems. In the preceding chapter, an information storage and retrieval format is defined to provide for effective and flexible manipulation of data related on a spatio-temporal basis. Algorithmic procedures are described for the manipulation and analysis of these aggregated data. A revised methodology is required for the development of these information bases from the newly available earth imagery.

Large quantities of significant resource information are embedded in earth imagery, which is available on a continuing basis, particularly through the NASA Earth Resources Technology satellite systems. Consequently, major image analysis and interpretation aids are implemented in IRIMAS to enable an interpreter to interactively enhance, classify, transform and display the image product. Additional software is provided to allow the analyst to save his interpretation in the format of the IRIMAS data files. These major

image analysis components provide the capability for the entry of image interpreted resource information into the complete IRIMAS system.

Image Analysis Functions

The implemented image analysis functions are ones which have been proven by repeated usage for resource survey and reconnaissance applications. Image enhancement by smoothing and edge enhancement techniques are established as useful procedures in the detection and delineation of specific image features [36]. The smoothing functions reduce the effects of random image noise on the interpretability of the data, while the edge enhancement functions provide a method whereby boundaries inherent in the image are enhanced and their detectability improved. These image enhancement techniques are particularly restricted to single spectral bands in imagery and do not offer the flexibility for the handling of multiband information (extensions to these multidimensional data sets, however, seem reasonable).

Multispectral imagery requires a more complicated set of procedures, specifically clustering or classification. These analysis techniques essentially compute similarity functions between data points and group data points which are similar into categories which are dissimilar from

other categories. The similarity functions may be either deterministically or stochastically derived.

An additional enhancement technique, density slicing, is beneficial in those cases where the magnitude resolution of the data is restricted either by system or interpretability constraints. This technique requires the ordering of the data into groups which contain equal numbers of elements. Assignment of values or symbols to each group completes the "density slicing" transformation and improves the ability to distinguish homogeneous categories by reducing the confusion generated by interpreting image elements with a wide range of values.

Smoothing

The purpose of the image smoothing function is to suppress or average the random noise introduced into the data by the measurement system and to reject and replace those image elements which appear to be improper with respect to the surrounding picture elements. The smoothing operation which is implemented is called noise-point elimination [4]. This is a "neighborhood" operation where several points in the vicinity of the element being tested are used in the smoothing operation. The noise-point elimination technique replaces a point by the average of its neighbors if the point value exceeds that average by

some preset threshold. This technique has the least effect on class boundaries while still providing some correction for image data noise. Additional approaches to the smoothing operation include replacing a specified point by the average of its neighbors, however, this approach tends to soften the edges in an image and thus tends to be detrimental to resource interpretation.

Edge and Line Enhancement

The enhancement of edges, lines and boundaries in an image is significant for improved interpretation. Broad categories can be brought into sharper distinction through application of line and edge enhancement algorithms. An item of caution in edge enhancement concerns the enhancement of lines and edges with regard to a particular direction. The results of some enhancement procedures are affected by the particular orientation of the edge features. A suitable enhancement algorithm which reduces the effect of this directional dependence is the gradient of the image function at the point being enhanced. In this operation, the gradient

$$f(x,y) = \left[\frac{\partial f(x,y)}{\partial x}^2 + \frac{\partial f(x,y)}{\partial y}^2 \right]^{1/2}$$

is computed to replace the data point. Computation of this function for digital pictures can be approximated by

$$P_{ij} = [(P_{i+1,j} - P_{i,j})^2 + (P_{i,j+1} - P_{i,j})^2]^{1/2},$$

where P_{ij} is the i -th picture element from the j -th row.

The gradient operation, and for that matter differentiation or differencing as a whole, directly relates to the concept of spatial filtering prevalent in image enhancement literature today [36]. Two dimensional spatial frequency domain transformations, such as the Fourier transform, are used to convert the image function from the spatial domain to the spatial frequency domain. Selection of the high frequency components in the spatial frequency domain and inverse transformation emphasizes or enhances those areas in the image where the image function is rapidly changing as a function of the coordinate variables. Thus, the enhancement of the image occurs at the boundaries or edges between categories of varying image densities. In an analogous manner, smoothing effectively selects and enhances the low spatial frequency components of the image, essentially providing low-pass filtering.

Density Slicing

The two-variable (dimension) image function which assigns a reflectance value to each coordinate point in the image plane, typically is quantized into numerous levels. The ERTS-1 MSS system for example quantized its image measurements into 64 distinct levels. The vari-

ability in scene radiance, signal to noise ratios, and quantization levels all affect the distribution of picture elements in the range of this discrete function. Manual image interpretation cannot effectively utilize this wide range of information values. Consequently, a transformation of the fine density resolution into a more suitable (coarser) range for manual image interpretation is appropriate. The most obvious procedure for effecting this transformation is to linearly modify the set quantization levels such that a smaller number of density values (e.g. 8) are produced for interpretation purposes.

This linear transformation is unacceptable due to the wide variability in the scene illumination and also in the reflectance of specific categories of interest. Typically, the image function for a specific scene may only cover 20-30% of the allowable quantization range. A linear transformation then assigns that percentage of the available quantization levels to the whole image, resulting in a significant loss of image information. A non-linear transformation seems appropriate.

The image density transformation which provides the greatest preservation of information, non-linearly sets new quantization levels according to the distribution of the image points in the range of the discrete image functions. These assignments are made so that equal

numbers of image points are assigned to each quantization level. Constructing the quantization levels in this manner allows each image density value to have an equal weighting with respect to every other image point.

Clustering

Clustering is normally thought of in terms of classification, since in general this category of techniques groups image points into classes where all the class members have similar image density values and non-members of a class are dissimilar. Numerous clustering algorithms have been described [6] and utilized as classification tools. The major drawback to the use of clustering as a classification mechanism is the large amount of processing required for multiple passes through the data in search of some suitable collection of categories with adequate classification accuracy. Additionally, the classes selected by a clustering algorithm generally do not assure that the classes determined solely from the spectral distribution of measured reflectance are those of concern to a resource manager. Consequently, the clustering techniques have been to a large extent shadowed by other classification techniques, for example, maximum likelihood. Clustering does offer specific advantages where the accumulation of similar points and the assignment of a

single grey scale to these groupings may be useful as an enhancement aid. IRIMAS uses a single pass clustering algorithm to group similar image points which lie together in a general area.

The clustering technique implemented [37] provides a reasonable trade-off between classification and speed of computation. The algorithm also uses a norm for the computation of point similarity which is more reasonable in terms of the actual classification assumptions than the maximum likelihood or a similarly derived "least squares" categorization technique. The distance function used for the clustering algorithm is the ℓ_∞ metric d ,

$$d = \max |x_i - x_j|$$

where the distance in the measurement space is the maximum of the absolute values of the differences between the spectral components. It was found in [37] that this distance measure provided an increased sensitivity to spectral change, suppressed the smearing of categories in the measurement space, and its success could be more readily predicted from individual data band characteristics than other measures averaging all the bands.

Spatial clustering is determined by computing the spectral cluster center and by adding points to the cluster where the distance from the point to the cluster center was less than a specified threshold distance away. The

clustering algorithm with its associated measure is implemented in IRIMAS to provide local enhancement of class associations.

Many resource categories of interest have elements which are aggregated spatially. The implemented clustering algorithm builds a table of clusters as it processes through the data. A new data point is compared to the cluster to which the previous point belonged, if the new data point is determined to not belong to that cluster the point is compared to the cluster of the point that lies near it on the preceding scan line. For typical cluster-classification, if this test fails, the point would be compared to all existing clusters before a new cluster was initiated. However, for the enhancement mode in which this procedure is utilized, the maintenance of a large number of clusters is generally not desirable. Therefore, if the point does not belong to the clusters of either of the two preceding points, a new cluster is initiated. At the end of each scan line, the clusters not used for that scan line may be reinitialized for use since the point aggregations which defined them are disjoint from the subsequent data points.

This cluster enhancement algorithm has the net effect of smoothing data within spatial aggregation while enhancing their boundaries with other groupings. Such a

procedure is helpful in identifying major data groupings particularly to aid in the selection of training sets for a supervised classification mode.

Classification

A supervised classification procedure is utilized by IRIMAS to provide class selection and identification from multispectral image function inputs. In a supervised classification mode, the classes to be identified are pre-specified to the algorithm which compares the new data to these specified classes and determines which class best describes the point. The classification procedure which has produced the best classification results operationally is the maximum likelihood classification procedure applied at the Laboratory for Applications of Remote Sensing, at Purdue University [28]. This algorithm has been generally accepted and versions of the LARSYS series of software packages are in wide usage. Consequently, the maximum likelihood classification procedure is implemented in IRIMAS.

The maximum likelihood classification algorithm assumes that the data fall into categories which may be described by multivariate Gaussian distributions with differing means and covariances. The implementation of the classifier algorithm computes the value of the con-

ditional probability density function for each category to be considered, and infers the membership of the data point to that category for which the conditional density function is the largest.

Class means and covariances are estimated from the measurement of signatures which are known or assumed to belong to that class. These estimated statistical parameters are then used in the classification algorithms. A class mean (for class i) is defined to be

$$\bar{X}_i = \frac{\sum X_i}{n}$$

and the estimated class covariance is

$$\Sigma_i = \frac{1}{n-1} \sum (X_i - \bar{X}_i) (X_i - \bar{X}_i)^T \quad \text{where}$$

X_i is the n -dimensional measurement signature for the i -th class. The conditional probability then becomes

$$f(X|\omega_i) = \frac{1}{(2\pi)^{n/2} |\Sigma_i|} \exp \left[-1/2 (X - \bar{X}_i)^T \Sigma_i^{-1} (X - \bar{X}_i) \right]$$

where X is the measurement signature, ω_i is the i -th class and \bar{X}_i , and Σ_i are the estimated i -th class means and covariances. The matrix operations $()^T$, $| \quad |$ and $()^{-1}$ correspond to transpose, the determinant, and the inverse of a given matrix. The classification decision for two classes then is

$$\omega = \begin{cases} \omega_0 & \text{if } f(x|\omega_0) \geq f(x|\omega_1) \\ \omega_1 & \text{if } f(x|\omega_1) < f(x|\omega_0) \end{cases}$$

for the data point X. An additional test is inserted into the classification algorithm. The possibility exists that a data point does not belong to any of the data classes specified. In this case, the data point is expected to have a low probability of belonging to any of the classes. This point is then assigned to the null or zero class where the largest value of the density function is determined to fall below a preset threshold. Using this rejection class minimizes the chance of identifying a measurement as belonging to a class when in fact the class to which it belongs was not specified.

Image Processing Software

A software component entitled PROCES is implemented in IRIMAS to provide the image processing functions or to reference specific subroutines supporting application of these analysis techniques. This software driver has two specific functions, the conversion of data byte strings into integer arrays for processing; and to effect the analysis function through either calls to external subroutines, or internal implementation of the analysis techniques.

The operation of the software driver PROCES is controlled by basically two parameters. The first parameter PASS indicates whether the driver should be in an

initialization, processing or termination mode. Entry into PROCES with PASS equal to 1 causes the driver to initialize and zero the appropriate arrays and variables. Entry with PASS equal to 0 indicates normal processing of the input data values, and entry with PASS equal -1 indicates that processing is concluded and termination of the analysis sequence is to be undertaken. The second control parameter identifies the specific analysis technique to be applied. In Table V-1, is shown the operation codes and the specific enhancement or classification technique to be applied.

In addition to the processing techniques described, the IRIMAS software system allows operator definition of a special transform to be applied to the data. This implementation utilizes the syntax drivers and interpretative subroutines previously described in Chapter IV for specification of data transformations. The capability provides a means for the application of specific discipline related transformation without major system modification. An operational flow chart for PROCES is shown in Figure V-1.

TABLE V-1

PROCES Operation Codes for Enhancement and Classification

<u>Code</u>	<u>Function</u>
0	No Operation
1	Classifier Training
2	Maximum Likelihood Classification
3	Algorithmic Clustering
4	Edge Enhancement
5	Smoothing
6	Histogram Preparation
7	Operator-defined Transformation

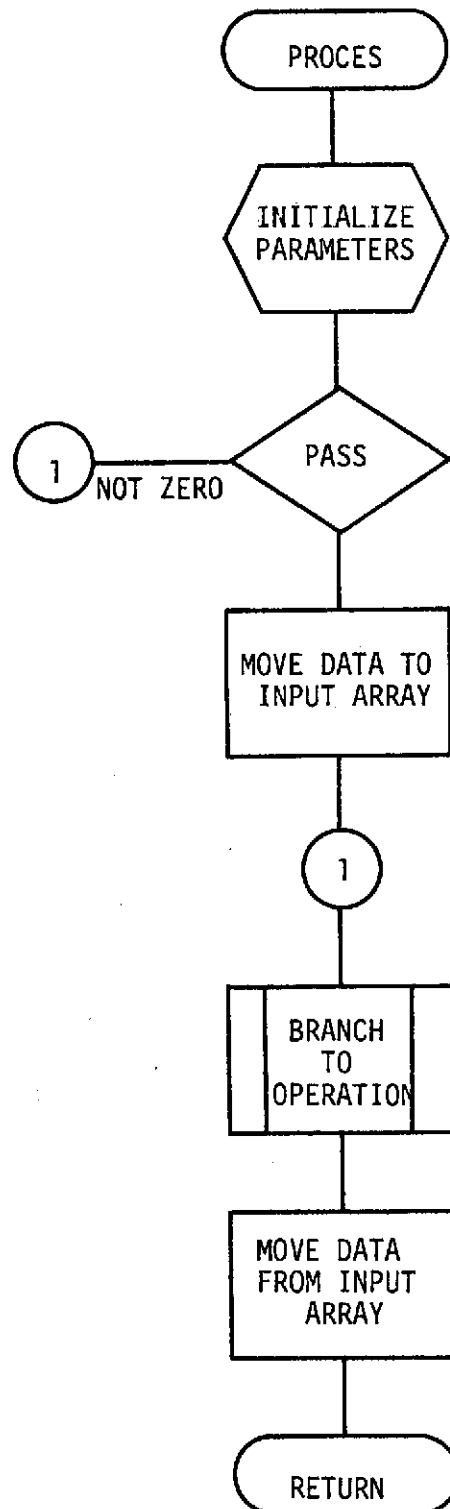


Figure V-1. Logical Flow Diagram for PROCES

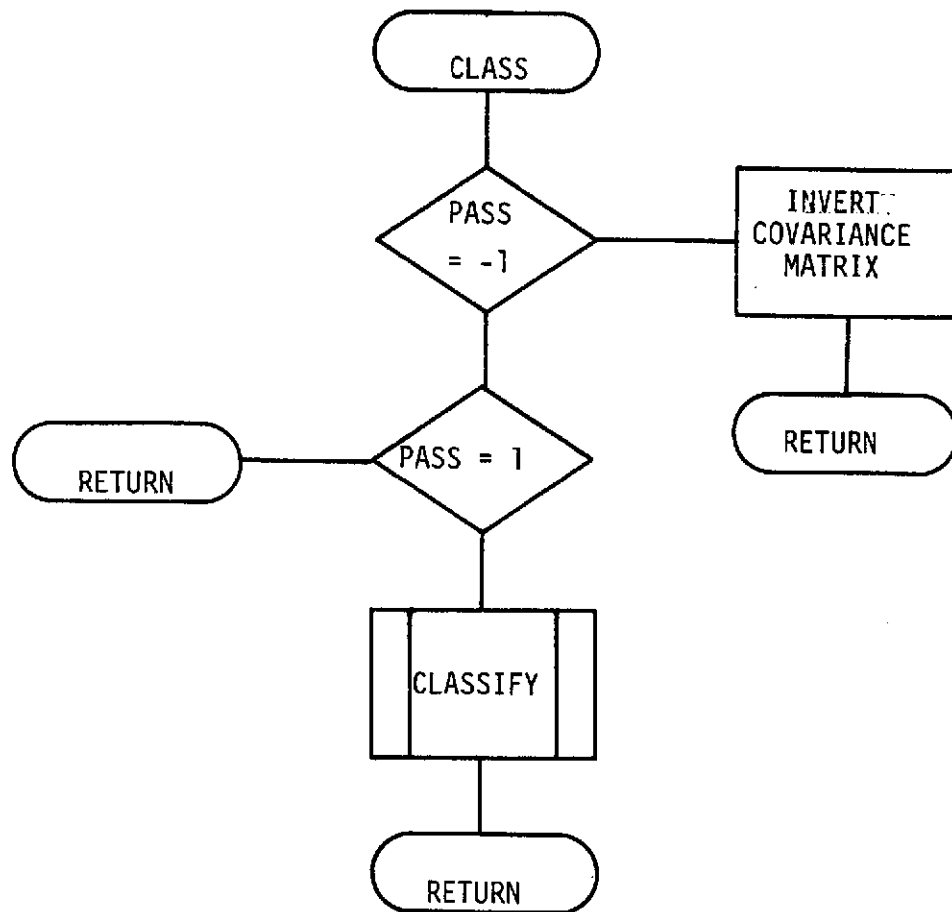


Figure V-1. Logical Flow Diagram for PROCES (cont.)

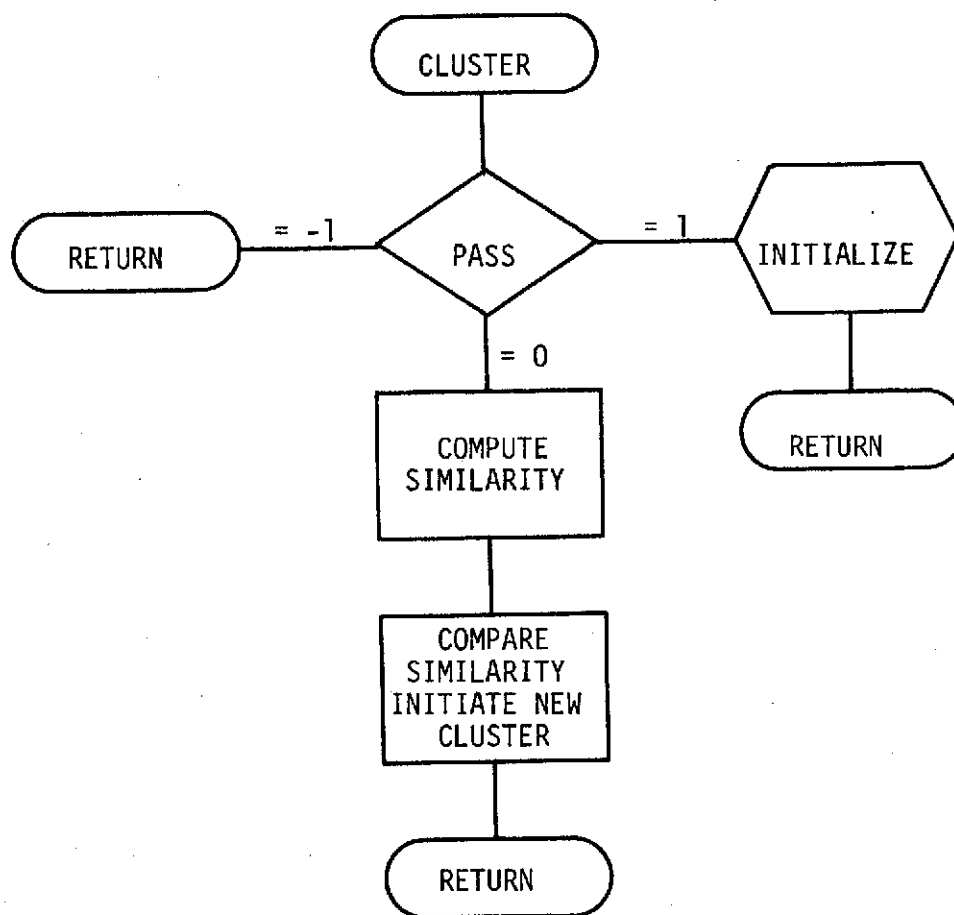


Figure V-1. Logical Flow Diagram for PROCES (cont.)

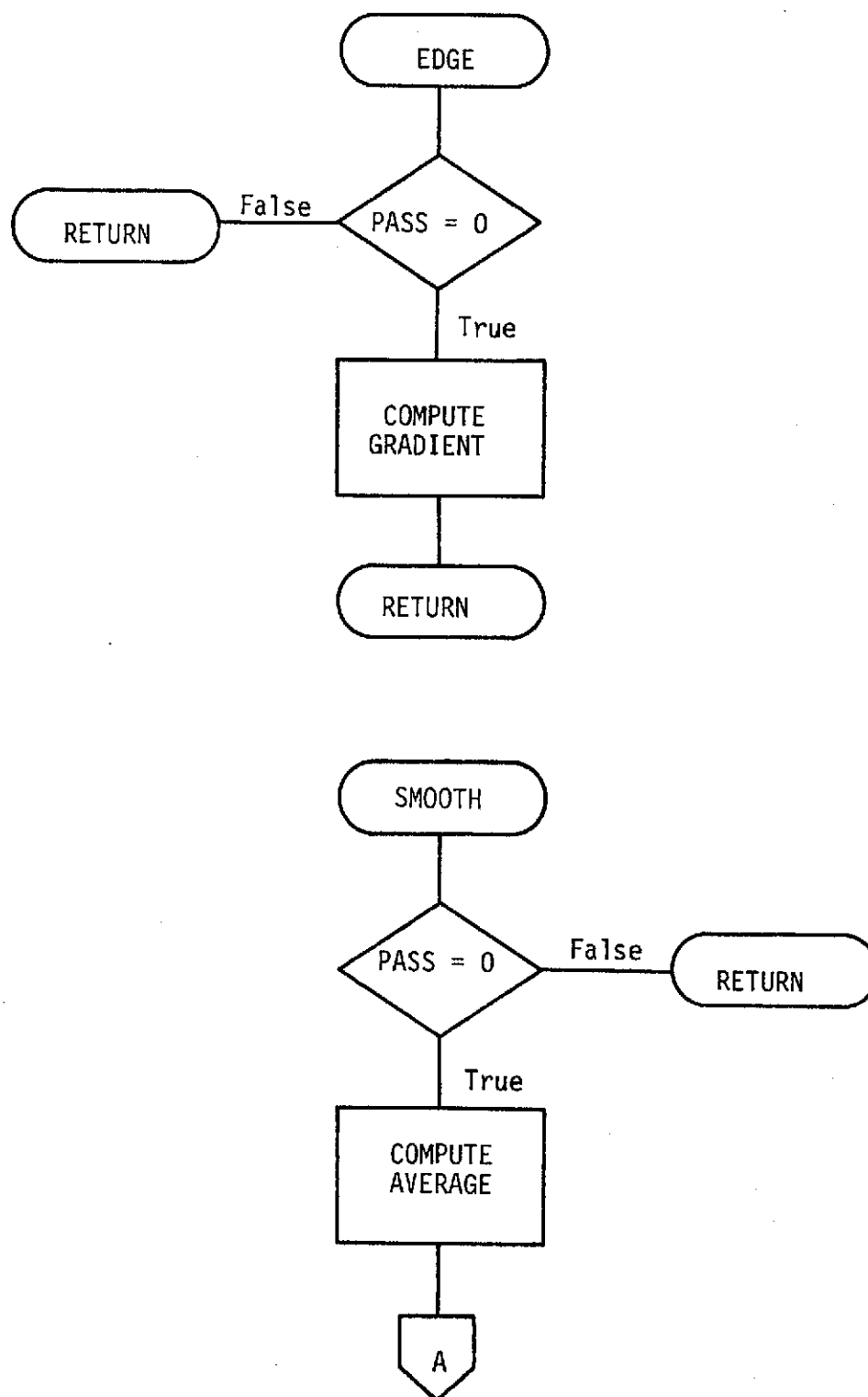


Figure V-1. Logical Flow Diagram for PROCES (cont.)

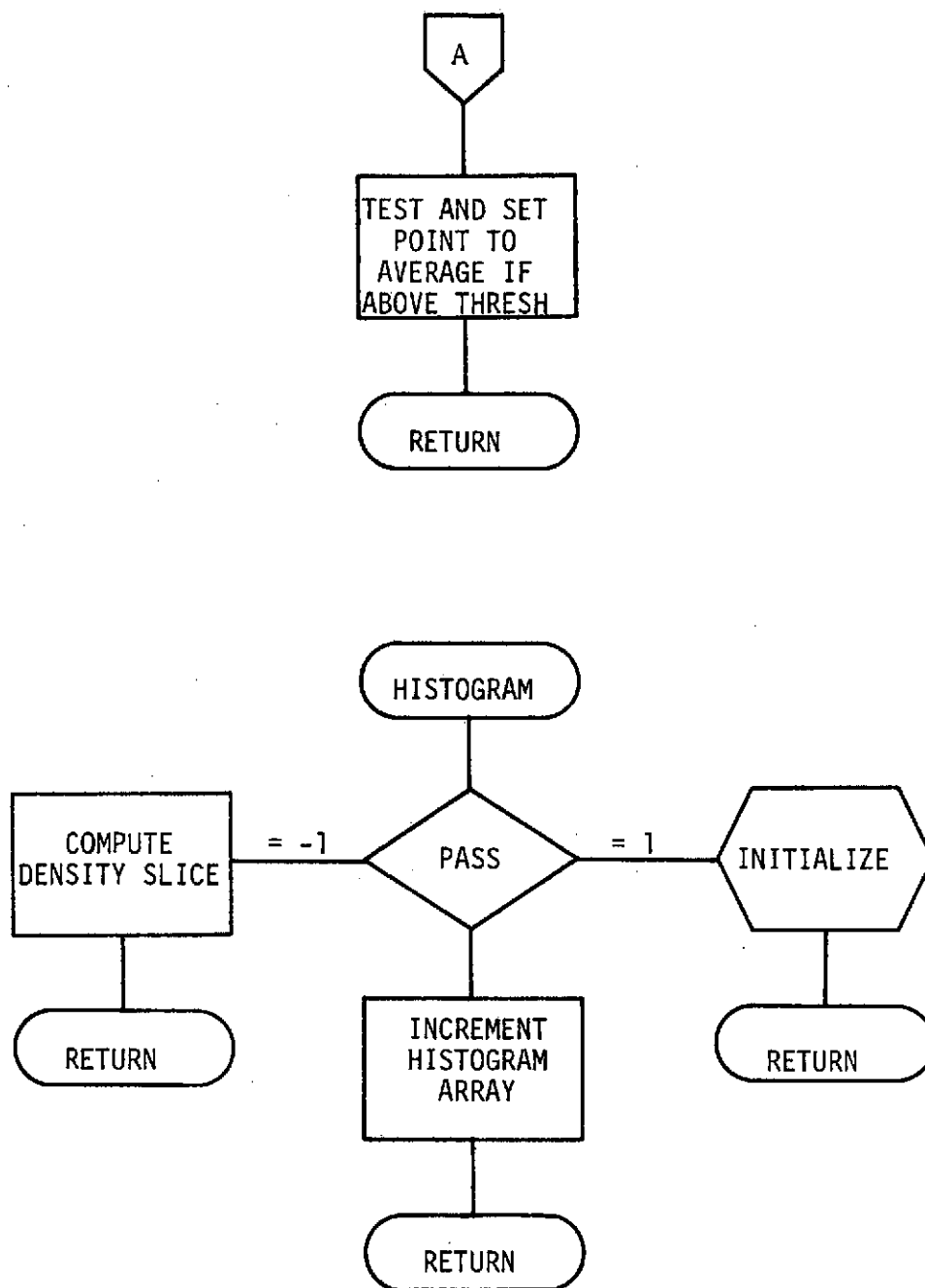


Figure V-1. Logical Flow Diagram for PROCES (cont.)

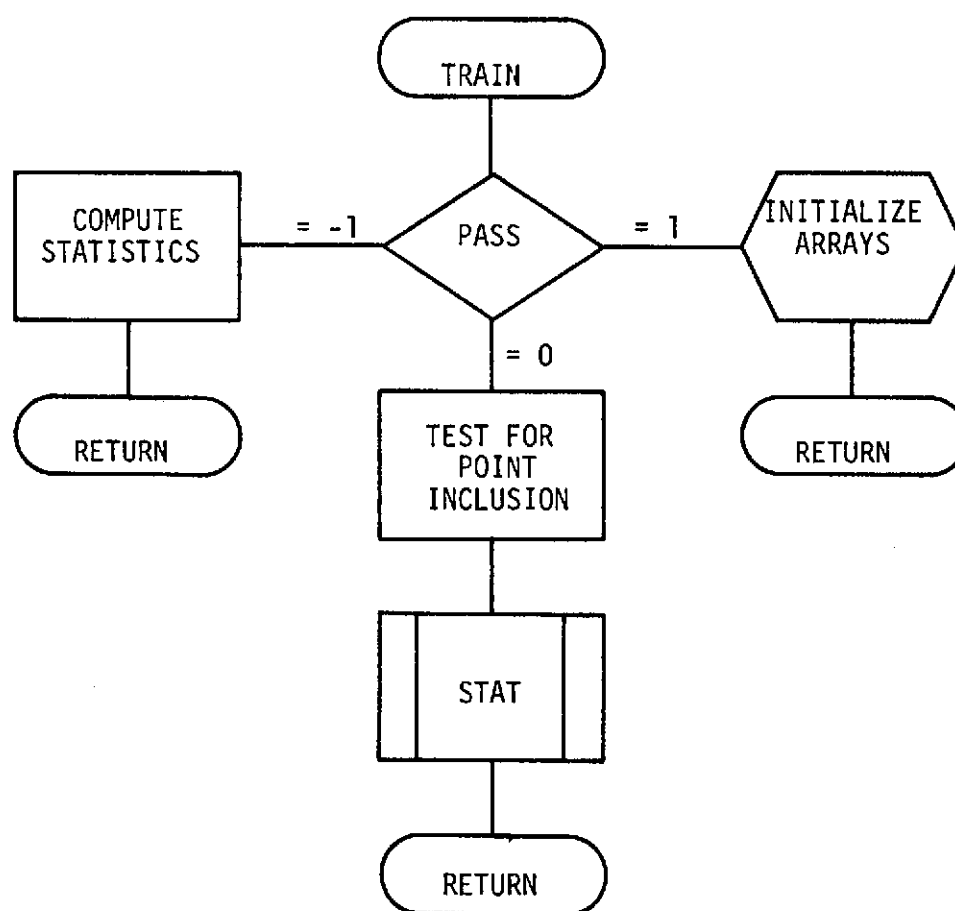


Figure V-1. Logical Flow Diagram for PROCES (cont.)

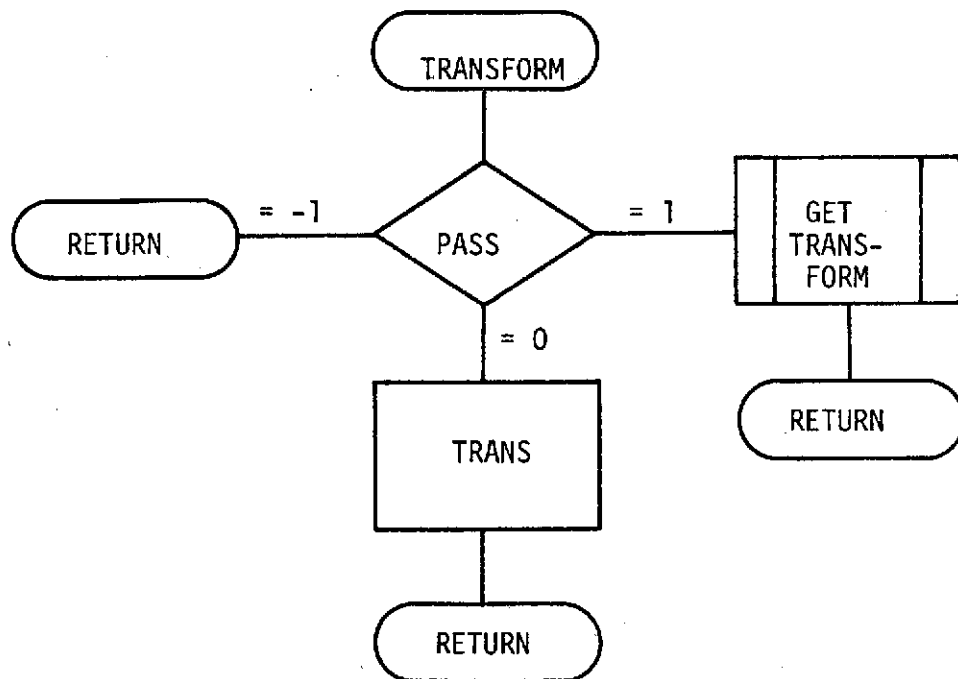


Figure V-1. Logical Flow Diagram for PROCES (cont.)

CHAPTER VI

SYSTEM OPERATION

The preceding chapters discuss the general requirements for a resource information management and analysis system to develop an effective interface between a resource planning function and the information available from remotely sensed data. An information format and techniques for information retrieved from data stored in that format are developed. Image analysis functions which enable an interpreter to analyze satellite imagery interpreting resource information, and to save it in data files for future reference and analysis are also discussed. Additionally, the software subroutine and techniques developed for these analyses and information data handling are described. This chapter concludes the detailed description of data handling and analysis techniques implemented, by explaining the integration of these individual subprograms into the Interactive Resource Information Management and Analysis System, IRIMAS, and by further defining the actual interface with the analyst/system operator.

System Control

The software system IRIMAS has three principal function areas: 1) input; 2) processing, and; 3) output, which are operator specified. The input function takes

data from an input data file and converts them into a usable internal format for IRIMAS. The processing function implements data and coordinate transformation, as well as image enhancement and classification algorithms. The output function provides image display, printer tabulation, and information format output to IRIMAS data files.

Operator interface has been accomplished in two primary ways. Operator directives are provided to the system for definition of the system tasks and are interpreted according to a syntax table in a manner similar to the procedures previously described for syntax analysis. The interpreted syntax provides operation codes which are saved for further interpretation by the system to define specific system operations related to data input and analysis. Specification of the output data product, or data display is also provided. Operation directives for control and termination of the system tasks are also implemented.

During the course of system analysis operations, operator inputs are also required to define analysis parameters. These are obtained by computer initiated queries to the system operator. The response to these questions provides the required information for the system analyses to continue. This approach to communication allows the computer to generate sufficient descriptive narrative accompanying each request to familiarize the

operator with the data input requested.

Operator directives for system task specifications define input, processing, output and control functions. These directives are saved in a stack array and are sequentially executed by the system. Additionally, the processing functions of IRIMAS are controlled by a loop parameter, PASS. This parameter specifies whether the system is in an initialization mode prior to the input of data (PASS=1), an operation or processing mode (PASS=0) where the data is available for processing, or a termination mode (PASS=-1) for the completion of processing at the end of the available data. The pass parameter is primarily controlled by the input functions which set the parameter according to the availability of data to the processing subsystem.

Analysis processing in IRIMAS is accomplished by each system element on single data entries which are provided by each call to an input driver. Consequently, a loop directive, "GO TO #", is installed to set up the processing loop back to the driving input function. The loop is broken when the input function detects an end-of-data status and sets the PASS parameter to the termination mode. At this time, operations subsequent to the input function are executed in the termination mode, however, the control branch does not occur in the termination mode. Operator

control directives PAUSE and STOP halt the execution of the specified functions. The major difference between these functions is that the system must be restarted after a STOP, and for a PAUSE the system resets to input additional operator function directives. The allowable operator directives are tabulated in Table VI-1, and are explained in following paragraphs.

The system IRIMAS is developed on a minicomputer system configured with printer and color CRT output displays. Restricted by the application of the display techniques, the system generates an output format with both an image and alphanumeric representation. It is anticipated that in general, the color CRT system is not readily available to all minicomputer facilities and therefore the output products have been designed for display on line printers. Several output display formats are provided by IRIMAS. These are primarily graphics or image output to either the printer or to the CRT, and tabulations of analyses parameter providing on the printer.

IRIMAS Input Drivers

The IRIMAS system accepts two basic varieties of input data. The first input variety is the standard IRIMAS information data format described in Chapter IV. In this format, data are saved as strings of characters specifying

TABLE VI-1
IRIMAS Operator Directives

<u>Directive</u>	<u>Function</u>
INPUT	Input from IRIMAS files
ERTS	Input from ERTS tapes
TRAIN	Classification Training
CLASSIFY	Maximum Likelihood Classification
CLUSTER	Algorithmic Clustering
EDGE	Edge Enhancement
SMOOTH	Smoothing
SLICE	Histogram Preparation
TRANSFORM	Interactive Image Data Transformation
XDATA	Interactive Information Data Transformation
XCOORD	Interactive Coordinate Transformation
EXPAND	Algorithmic Expansion of Information
DISPLAY	Output to CRT
MAP	Output to Printer
DEFINE	Set up IRIMAS Files
HISTOGRAM	Print Histogram
STATISTICS	Print Classification Parameters
DENSITY	Print Density Slice Levels
OUTPUT	Print Output Variables
STOP	Halt IRIMAS Processing
PAUSE	Pause in IRIMAS Processing

TABLE VI-1 (cont.)

GO TO #	Transfer to Statement Indicated
---------	---------------------------------

resource boundaries by coordinate locations, attribute values, and data set descriptions. It is the format which allows the interpretation of images to be uniquely saved and referenced without the intermediate interpretation steps. This format is also convenient for the inclusion of additional resource information obtained from ground survey or other techniques, provided the information can be referenced to a common coordinate basis. The software driver INPUT implements this function.

The second input source supported by IRIMAS is from the Earth Resources Technology Satellite. This data input is from digital magnetic tape and is in a fixed raster scan format. It contains the four spectral bands of reflectance data measured by the ERTS satellite multispectral scanner system. The IRIMAS component "ERTS" provides the software support for the reading of ERTS digital data, and converting the ERTS format into an internal format for processing.

INPUT

The software driver INPUT implemented in IRIMAS facilitates resource data input from IRIMAS data files. The information format is defined by a syntax table, and syntax interpretative techniques are utilized to convert information from the data files into a machine internal

format. Fundamental to the system is the internal buffer area where the information from the input data file is kept. This buffer area is named INFORM and includes three specific arrays. The first array, named ATTR, holds the values of the attributes which refer to the aggregate of information elements being processed from the input file; the second array, COORD, holds the x-y coordinate boundary of the category; and the array NAME holds the name of the data entry. This internal format is fixed to allow effective processing of the referenced information by processing subroutines.

ERTS

Input to the IRIMAS processing system from the Earth Resources Technology Satellite digital image tapes is accomplished by the software driver ERTS. This subprogram allows operator selection of the tape line and data cell coordinates of the ERTS tape and also the selection of either one of the four ERTS bands from the tape or all of the bands. After the selection of the ERTS input parameters, the subprogram reads in the ERTS data, a line at a time, converts the data according to the bands selected and stores the data into a raster buffer area named BUFFER. This buffer array holds two lines of ERTS data simultaneously to allow neighborhood enhancement processing.

On the initial entry to the input driver, ERTS tape line, data cell and band parameter specifications are requested from the operator. The subprogram keeps a count of the number of tape lines read from the tape and either skips or backspaces to the line requested before processing begins. Successive calls to ERTS cause a fetch of an additional data raster line and provides for processing into the buffer area. After 128 lines of ERTS data are processed, the driver interrupts the processing loop with an end-of-data indication which sets the PASS parameter for the termination mode.

Processing

The second major set of functions implemented in IRIMAS are concerned with the processing of the input data. The processing drivers deal with one of three data storage arrays. The first two arrays, INFORM and BUFFER, were previously described (Chapters IV and V) as internal format storage for data input from IRIMAS data files and from Earth Resources Technology Satellite digital tapes, respectively. The third array area, FRAME, is used as an intermediate storage buffer for grid format data and corresponds to an output display window. Information inputs from IRIMAS data files may be expanded into a grid format and saved in this storage array. Different infor-

mation entries may be overlapped in the array and analysis performed. Additionally, the expansion into the frame array provides a convenient format from which data may be displayed.

Processing of IRIMAS input data is interactively specified by an operator. The input commands to the system together with their meanings and referenced processing drivers are described in subsequent paragraphs.

TRAIN

The TRAIN operation allows the computation of mean value and covariance matrix estimates which are used in the maximum likelihood algorithm. Initialization procedures for this operation request information from the system operator to set the number of classes to be specified, and the names and rectangular boundary coordinates for each class. The coordinate of each data point during the processing mode are compared to the coordinates specified for the class boundaries. For those points lying within the boundaries, the values for each of band specified by ERTS input processing are accumulated into buffers in preparation for the application of the statistics estimation formula. Entry into this software processing routine during the termination pass causes the statistics estimates to be completed and the class statistics to be saved.

CLASSIFY

This classification directive specifies maximum likelihood classification of the input data. During the initialization mode, the covariance statistics computed by the TRAIN processor are inverted and their determinant calculated for classification processing. During the process mode, each input data point is utilized by the classification processor to determine its most likely class and to compare the probability of its being produced by that class within a preset threshold. If the probability of the point having occurred, (conditioned on the most likely class), exceeds the threshold; the first point attribute, or band value, is set equal to the class number, otherwise, it is set equal to zero. The number of the bands utilized in the classification mode is determined by the bands specified in the ERTS input driver. Termination processing has no effect on the classification procedure.

CLUSTER

The system directive CLUSTER provides for the clustering of adjacent picture elements whose computed similarity is below a preset threshold. Initialization causes each set of cluster parameters to be zeroed. Note that the cluster parameters are saved in the same buffer area as the training set boundaries so that the two

processing techniques may not be applied simultaneously. The processing mode of the cluster algorithm computes the clusters to which incoming data vectors belong and output the cluster number in the position of the first input data band. The termination mode does not affect the cluster algorithm and the number input data bands processed is set by the ERTS routine.

EDGE

The system directive EDGE specifies edge enhancement of the input data according to the gradient technique described in Chapter V. A neighborhood of three is utilized, consisting of the data point and adjacent neighbors which occur later in the input data string. Single band data are assumed and the edge enhancement operation is not affected by initializing or termination modes.

SMOOTH

The system directive for smoothing provides for noise-point elimination by comparing a data point with the average of its four-point neighborhood. If the data point is significantly different, exceeding a preset threshold, the data point is replaced by the neighborhood average. A single band is assumed, and initialization and termina-

tion modes are ignored.

SLICE

The computation of a 128 value-level histogram is specified by the SLICE directive. The break points for density slicing the data values into equal number bins are also computed. Initialization causes the histogram data arrays to be zeroed. During the processing mode, the array elements are incremented according to the value of the data point processed. Only one data band is assumed. During the termination mode, the density slice breakpoint levels are calculated.

TRANSFORM

The system directive TRANSFORM sets up interactive programming of input data transformations. Upon initialization, the processor requests a definition of the transforms to be utilized. The input statements are checked against the data transformation syntax described by SNTBL2 and interpreted by SUBR.2 previously described. The transformation definition terminates with the input of the "END" statement. The interpretation of the defined transform is saved and during the processing mode is utilized by subroutine TRANS to effect the data transformation. The number of bands available for processing depends upon

the specification provided by ERTS. The bands are referenced as attribute variables (A1, A2, etc.). The termination mode is ignored by the transformation processor.

XDATA

This operation specified by XDATA provides for the transformation of data input from IRIMAS data files into the INFORM buffer arrays. During the initialization procedure, the driver calls for operator specification of the data transformations according to the data transformation syntax defined by SNTBL2 and interpreted by the subroutine SUBR.2. During the processing mode, the input data are transformed in accordance with the defined transformation string interpreted by the subroutine TRANS. The termination mode does not affect the XDATA driver.

XCOORD

The processing directive, XCOORD, provides for the coordinate transformation of data input from the IRIMAS data files into the storage array INFORM. Upon initialization, the driver requests the coordinate transformation definition from the operator and sets up the transformation string to be interpreted by TRANS during the processing phase. During processing, the coordinates are unstacked

from the INFORM coordinate stack and are transformed by the interpreter TRANS. Upon completion, the transformed coordinates are returned to the INFORM coordinate stack. The termination mode does affect the XCOORD processor.

EXPAND

This processing operation specifies the key routine in the processing of IRIMAS data files. The EXPAND software implements the clipping, line generation, and area generation algorithms described in Chapter IV for the conversion of IRIMAS information files into grid format data. These grid format data are saved in the storage array FRAME, and interfaced directly with the display peripherals. The processor EXPAND ignores the initialization and termination modes.

System Output

The final set of major functions implemented in IRIMAS provide for the output and display of the processed data. Output processing components together with the numonics utilized for operator specifications are described.

DISPLAY

The output directive DISPLAY provides for the CRT (cathode ray tube) display image data lines maintained

in the BUFFER array. These image lines are displayed on the system dynamic color display [38], currently in a 128 x 128 picture element array. This output processor is typically specified within an input-processing loop since only single lines of data are output. Neither initialization nor termination processing modes affect this processor.

MAP

The output directive MAP provides for the output of image data to a printer device. Density slicing of the data and assignment to alphanumeric symbols according to parameters provided by the SLICE processing provide a hard copy of the data image. This processor nominally provides a 64-column image display however, larger line lengths are printed in multiple parts. The processor is unaffected by the initialization or termination modes.

DEFINE

The output element DEFINE is a key module implemented in IRIMAS and provides for the interactive specification and output of IRIMAS data files. The primary elements for the construction of an IRIMAS data file are requested from the operators. These data are imbedded into the sentences prescribed by the data file syntax and outputs these

sentences as records to an output file. This module does not recognize any of the processing modes.

HISTOGRAM, STATISTICS, DENSITY, OUTPUT

Implemented output modules provide printed hard copy output from the analysis operations described. The directive HISTOGRAM, and DENSITY specify the printing of a graphic display of the data histogram compiled by the processing module SLICE, and for the output of the break points computed by that processing function. The directive STATISTICS specifies output of the data means and covariances computed for the data classes defined in TRAIN, while the directive OUTPUT calls for the printing of the data contained in the output data arrays referenced by operator defined data transforms. The referenced output routines do not recognize the different processing modes.

Control

Three control directives are provided for operator control of processing. These directives include STOP, PAUSE and GO TO #. The STOP directive causes the program to terminate when this directive is encountered. This requires restarting of the IRIMAS system, which initializes certain processing variables, and zeros data arrays. The restart of IRIMAS also rewinds data input and output files.

For a temporary interruption in data processing, the operator directive PAUSE should be specified. This directive causes termination of the processing sequence when it is encountered, however, the interruption provide for a start operator definition upon system processing directives, as opposed to the complete restart required by STOP.

Since the input drivers and the processing modules handle a single data record at a time, the operator directive GO TO # has been established to set up the loop. This directive is used to loop back to the statement number "#", generally the input directive for multiple record processing. The GO TO # driver resets the processing mode from "initialization" to "processing" and branches for these two modes, however it is processing mode 0, a no operation function in the termination mode, which is set on completion of the input of data required.

Chapter VII describes by example, the operation and operator interface of the software system IRIMAS. Operator information statements and diagnostic messages are also discussed.

CHAPTER VII

SYSTEM DEMONSTRATION

The general specifications and specific implementation of the Interactive Resource Information Management and Analysis System, IRIMAS, were discussed in detail in previous chapters. The system is implemented in two parts, one dealing with data stored in the IRIMAS format, and the second dealing with the image analysis functions. This dual implementation is required by the memory size constraints of the current configuration of the Texas Instruments 980A minicomputer system on which the software was developed. This chapter shows by demonstrated example the features of the system, the interactive operator interface, and the diagnostic and error messages implemented.

The operator interface previously described is implemented by computer-initiated dialogue and operator response. The system requires operator definition of analysis functions and input parameters during phases of specific function execution. Diagnostic phrases and system error messages are output as required to provide operator assurance of orderly data processing and to point out procedural errors in processing.

The system is demonstrated in a manner consistent

with the two-part implementation. One demonstration deals with the information format processing and a second demonstration with the image processing function. A third demonstration on the use of the IRIMAS system in change monitoring is also included.

Information Processing

The portion of IRIMAS implemented to handle and analyze data in the IRIMAS data format is called INFORM. This program segment contains the syntax drivers for interpretation of data file information, and data and coordinate transformations. Also included is the software described in Chapter IV required to expand the boundary format representation into grid format representation. The output functions for CRT display and line printer mapping are also included. Data input is restricted to IRIMAS format data files and the software for creating these files is also implemented. This segment of the demonstration is intended to show the generation of an IRIMAS data file describing the well-known block TAM insignia of Texas A&M University. This insignia is used to represent resource information which might be specified in terms of the IRIMAS data files. It is recognized that this information relates to either points, lines, or areas. Subsequent productions of the IRIMAS system demonstrate

the reconstruction and display of the "resource information", as well as coordinate transformations which may be applied so that information from different sources may be aligned. Resource information such as the major highway systems in Brazos County may be similarly entered into the system and analyzed. Also shown in this demonstration are the operator directive sequences required for specification of these analyses.

Operator Directives

IRIMAS operator directives are required to specify the system functions. In Figure VII-1 is shown an example of the computer dialogue for several of the control directives. A syntax interpreter is utilized by the software system to implement system function directives.

The first entry demonstrated is an incorrect directive to the system. A syntax error is detected and the operator is instructed to re-enter the numbered line. The corrected line is entered and the software steps to the second line for directive entry.

In the second line is entered the symbol "@" which instructs the system that the operator desires to re-start the function definition phase. At this point the system resets, and re-initiates the definition phase.

The PAUSE directive and the STOP signify termination

```
//EXECUTE.  
IRIMAS--AN INTERACTIVE RESOURCE  
INFORMATION AND ANALYSIS SYSTEM  
ENTER ANALYSIS FUNCTIONS  
0001 GOTO 2  
      ^SYNTAX ERROR IN ENTRY--RE-ENTER  
0001 GO TO 1  
0002 ?  
ENTER ANALYSIS FUNCTIONS  
0001 PAUSE  
DEFINED FUNCTIONS ARE COMPLETED♦♦CONT.  
ENTER ANALYSIS FUNCTIONS  
0001 STOP  
NORMAL TERMINATION
```

Figure VII-1. Example of Operator Dialogue

**ORIGINAL PAGE IS
OF POOR QUALITY**

of the function definition phase and indicate the function sequence termination point. The difference in the two directives is that STOP causes termination of the system execution. The directive PAUSE signifies that the function definition is completed, however, the data contained in system buffer areas should be retained as further processing may be required. The system, upon completion of a PAUSE directive, indicates that the defined functions have been completed and that additional function definition is required.

Completion of a STOP directive causes an operator diagnostic to be printed to indicate normal system termination and halts the IRIMAS execution. These control functions are also demonstrated in Figure VII-1.

Data File Definition

The software subsystem DEFINE is developed to allow interactive operator definition of IRIMAS data files. This driver requests data information from the operator, formats the information into that specified by IRIMAS and data files. This driver requests data information from the operator, formats the information into that specified by IRIMAS and writes the data to an output file.

In Figure VII-2 is demonstrated the procedure for definition of IRIMAS data files. In this case, the opera-

IRIMAS--AN INTERACTIVE RESOURCE
INFORMATION AND ANALYSIS SYSTEM
ENTER ANALYSIS FUNCTIONS

Step 1 0001 DEFINE
0002 PAUSE

Step 2 INFORMATION DATA SET DEFINITION
TERMINATE ITEM ENTRIES WITH
EMPTY LINE.

Step 3 TERMINATE DATA SET WITH /♦
IDENTIFICATION(NAME,COORDINATES,
DESCRIPTIONS) DELIMIT ITEMS WITH ", "
♦♦♦TAM DEMONSTRATION♦♦♦

Step 4 ATTRIBUTE DESCRIPTIONS
ENTER ATTRIBUTE NAME: DESCRIPTION
NAME ATTRIBUTES A,B,C,ETC.
A:DEMONSTRATION=15

Step 5 DATA ENTRY DEFINITION
TYPE(A,P,L): P
NAME: MMMM
LIST ATTRIBUTE AND VALUE
ATTR: A
VALUE: 15
ATTR:
COORDINATE DEFINITION

X= 30
Y= 40

X= 30
Y= 20

X= 35
Y= 40

X= 40
Y= 20

X= 40
Y= 40

X=

Step 6 DO YOU WISH TO CONTINUE?N
DEFINED FUNCTIONS ARE COMPLETED♦♦CONT.

Figure VII-2. Procedure for Definition
of IRIMAS Data Files

ORIGINAL PAGE IS
OF POOR QUALITY

tor (Step 1) specifies the system function DEFINE, and conclusion of the system analysis function definition by the PAUSE statement.

Execution of the DEFINE function prints the explicative message which states that information data set definition is being accomplished (Step 2). Additionally, it is explained that individual data entries are terminated by the entry of a blank line, and the entire data set may be terminated by an end-of-file symbol, "/*". The definition of a data entry class may be restarted by the symbol "@".

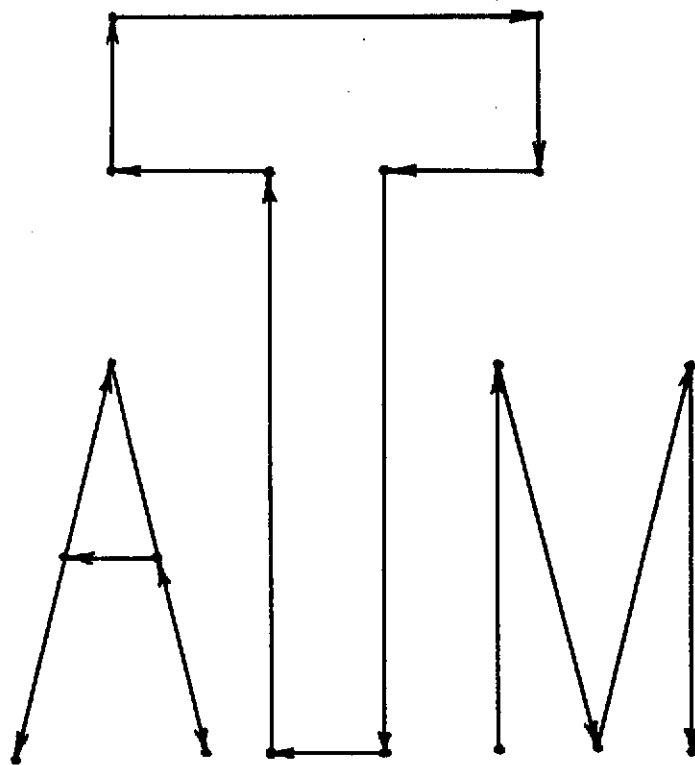
In Step 3, the data set identification is requested. This entry normally requires identification of the information to be entered, its source and degree of processing, and appropriate reference coordinates. In the example, the test identification, "****TAM DEMONSTRATION****", is entered. This step is terminated by an empty line. Step 4 calls for the entry of attribute descriptions to further define quantitative values utilized in data entry. As an example, the entry "A: Demonstration = 15", indicates that the attribute "A" has a value of 15 for the demonstration. These definitions are terminated by an empty line.

Step 5 begins the entry of quantitative data into the information system and the system calls for data entry definition. The first piece of information requested is the data entry type which is either point (P), line (L),

or area (A) information. In the example, point information is defined. The name of the data entry is requested next and a name "MMMM" is entered. Attribute names and values are requested from the operator with the attribute name first, and then the attribute value next. The attribute "A" is given a value of 15 in the example and attribute definition is terminated by the entry of an empty line. Next is requested the definition of data coordinates. At the conclusion of coordinate definition, an empty line is entered, and the operator is questioned as to whether he wishes to continue definition of additional data entries (Step 6). The response "N" terminates the define phase and completes the defined functions, whereas "Y" continues the process.

Data Reconstruction

An example of system operation is presented to demonstrate the handling of information data file point, line and area type of data. In Figure VII-3 is shown a traditional insignia of Texas A&M University and the coordinates of the lines constructing the letters are tabulated. These coordinates are used in the system example and the letter T is given a value of 15, A has a value of 12 and M has a value of 4. In Figure VII-4 is shown the operator input sequence to input, transform,



		Coordinates			
		T		A	
x	y			x	y
10	2			5	40
32	2			10	20
32	10			15	40
24	10			12	30
24	40			8	30
18	40			5	40
18	0				
10	10				
10	10				
10	2				
				M	
x	y			x	y
30	40			30	40
30	20			30	20
35	40			35	40
40	20			40	20
40	40			40	40

Figure VII-3. Traditional Insignia of Texas A&M University

```
♦READY♦  
//EXECUTE.  
IRIMAS--AN INTERACTIVE RESOURCE  
INFORMATION AND ANALYSIS SYSTEM  
ENTER ANALYSIS FUNCTIONS  
0001 INPUT  
0002 XDATA  
0003 EXPAND  
0004 GO TO 1  
0005 DISPLAY  
0006 MAP  
0007 STOP  
ENTER DATA TRANSFORMATIONS  
0001 Q0=A1  
0002 END
```

ORIGINAL PAGE IS
OF POOR QUALITY

Figure VII-4. Operator Input Sequence for
Computer Control

expand, and display these data on a computer generated printer map.

The system directive XDATA allows operator specification of transformation of input data points. The output variable " $\phi 0$ " is displayed, therefore the first attribute "A" is equated to " $\phi 0$ ". In the example, the entry type, which corresponds to "A0", is overridden and the END statement terminates the transformation definition. Input data entry types are overridden to produce point, line and area representations of the three letters. These printer maps are shown in Figures VII-5, 6, and 7.

Additionally, the system IRIMAS provides for coordinate translation, rotation and scaling through the XCOORD directive. In Figure VII-8 is shown the operator definition sequence for the translation and 45-degree rotation of the university insignia. The result of these operations is shown in Figure VII-9.

Brazos County Highway System

The major highways in Brazos County, Texas represent a transportation resource which might be used in an analysis of that county. In Figure VII-10 is shown a general highway map of Brazos County. Highway systems represented on Hwy. 6, Hwy. 6 business, Hwy. 21, Hwy. 30, Hwy. 60, and Hwy. 1179 are entered into an IRIMAS data file, and

0000							
0001							
0002		□			□		
0003							
0004							
0005							
0006							
0007							
0008							
0009							
0010		□		□	□		□
0011							
0012							
0013							
0014							
0015							
0016							
0017							
0018							
0019							
0020		L			D		D
0021							
0022							
0023							
0024							
0025							
0026							
0027							
0028							
0029							
0030		L	L				
0031							
0032							
0033							
0034							
0035							
0036							
0037							
0038							
0039							
0040		L		L	□	□	D D D D

Figure VII-5. TAM Point Representation

```

0000
0001
0002      000000000000000000000000
0003      0                                0
0004      0                                0
0005      0                                0
0006      0                                0
0007      0                                0
0008      0                                0
0009      0                                0
0010      000000000      000000000
0011                    0      0
0012                    0      0
0013                    0      0
0014                    0      0
0015                    0      0
0016                    0      0
0017                    0      0
0018                    0      0
0019                    0      0
0020          LL      0      0      D      DD
0021          LL      0      0      D      DD
0022          LL      0      0      D      DD
0023          LL      0      0      D      DD
0024          L L      0      0      D      D D
0025          L L      0      0      DD      D D
0026          L L      0      0      DD      D D
0027          L L      0      0      DD      D D
0028          L L      0      0      DD      D D
0029          L L      0      0      D D      D D
0030          LLLLLL      0      0      D D      D D
0031          L L      0      0      D D      D D
0032          LL L      0      0      D D      D D
0033          L L      0      0      D D      D D
0034          L L      0      0      D D      D D
0035          L L      0      0      D D      D D
0036          L LL      0      0      D D      D D
0037          L L      0      0      D DD      D
0038          L L      0      0      D DD      D
0039          L LL      0      0      D DD      D
0040          L L      00000000      D      D      D
0041

```

Figure VII-6. TAM Line Representation

```

0000
0001
0002
0003      000000000000000000000000
0004      000000000000000000000000
0005      000000000000000000000000
0006      000000000000000000000000
0007      000000000000000000000000
0008      000000000000000000000000
0009      000000000000000000000000
0010      0      0000000000000000
0011      00000000
0012      00000000
0013      00000000
0014      00000000
0015      00000000
0016      00000000
0017      00000000
0018      00000000
0019      00000000
0020      00000000
0021      LL      00000000      DD      D
0022      LL      00000000      DD      D
0023      LL      00000000      DD      D
0024      LLL     00000000      DD      DD
0025      LLLL    00000000      DDD     DD
0026      LLLL    00000000      DDD     DD
0027      LLLL    00000000      DDD     DD
0028      LLLLL   00000000      DDD     DDD
0029      LLLLLL  00000000      DDDD    DDD
0030      LLLLL   00000000      DDDD    DDD
0031      LL      LL      00000000      DDDD    DDD
0032      L       LL      00000000      DDDD    DDDD
0033      L       LLL     00000000      DDDD    DDDD
0034      LL      LL      00000000      DDDD    DDDD
0035      LL      LL      00000000      DDDD    DDDD
0036      L       LL      00000000      DDDD    DDDD
0037      LL      LL      00000000      DDDDDDDDDDD
0038      LL      LL      00000000      DDDDDDDDDDD
0039      LL      LL      00000000      DDDDDDDDDDD
0040      L       L      0      DDDDDDDDDDD
0041

```

Figure VII-7. TAM Area Representation

```
♦READY♦
//EXECUTE.
IRIMAS--AN INTERACTIVE RESOURCE
INFORMATION AND ANALYSIS SYSTEM
ENTER ANALYSIS FUNCTIONS
0001  INPUT
0002  XCOORD
0003  EXPAND
0004  GO TO 1
0005  DISPLAY
0006  MAP
0007  STOP
ENTER COORDINATE TRANSFORMATIONS
0001  TR(-20,-20)
0002  ROT(45)
0003  TR(20,20)
0004  END
$++++TAM++++//
&A:=15//
END OF DATA SET ON INPUT UNIT♦♦CONTINUE
```

Figure VII-8. Operator Definition for Rotation

```

0000
0001
0002
0003
0004
0005
0006
0007
0008      DDDD
0009      D  DD
0010      D  DD      DD
0011      DDD      DDD      DDD
0012      D      DD      DD DDD
0013      D      DD      DD DDD
0014      D      DD      DD DDD
0015      DDD      DD      DDDD      DD DDD
0016      D      DD      DDDDDD      DD D
0017      D      DDDD      DD      DD DDDDD DD
0018      D      D      DD      DDD      DD DDDDDD
0019      DD      D      DD      DD      DDD      D
0020      DD      D      DD      DD      DD      DDD
0021      DDDD      DD      DD      DD      DDD
0022      D      DD      DD      DD      D
0023      DDD      DD      DD
0024      DD DDD      DD      DD
0025      DDD      DD      DDD
0026      DD DDD      DD D
0027      DD DDDDD      DDDD
0028      DD D      DDDD      D
0029      DDD      DDDD
0030      DD      D
0031      DD
0032      DD
0033      DD
0034      D
0035
0036
♦READY♦

```

Figure VII-9. TAM 45-Degree Rotation

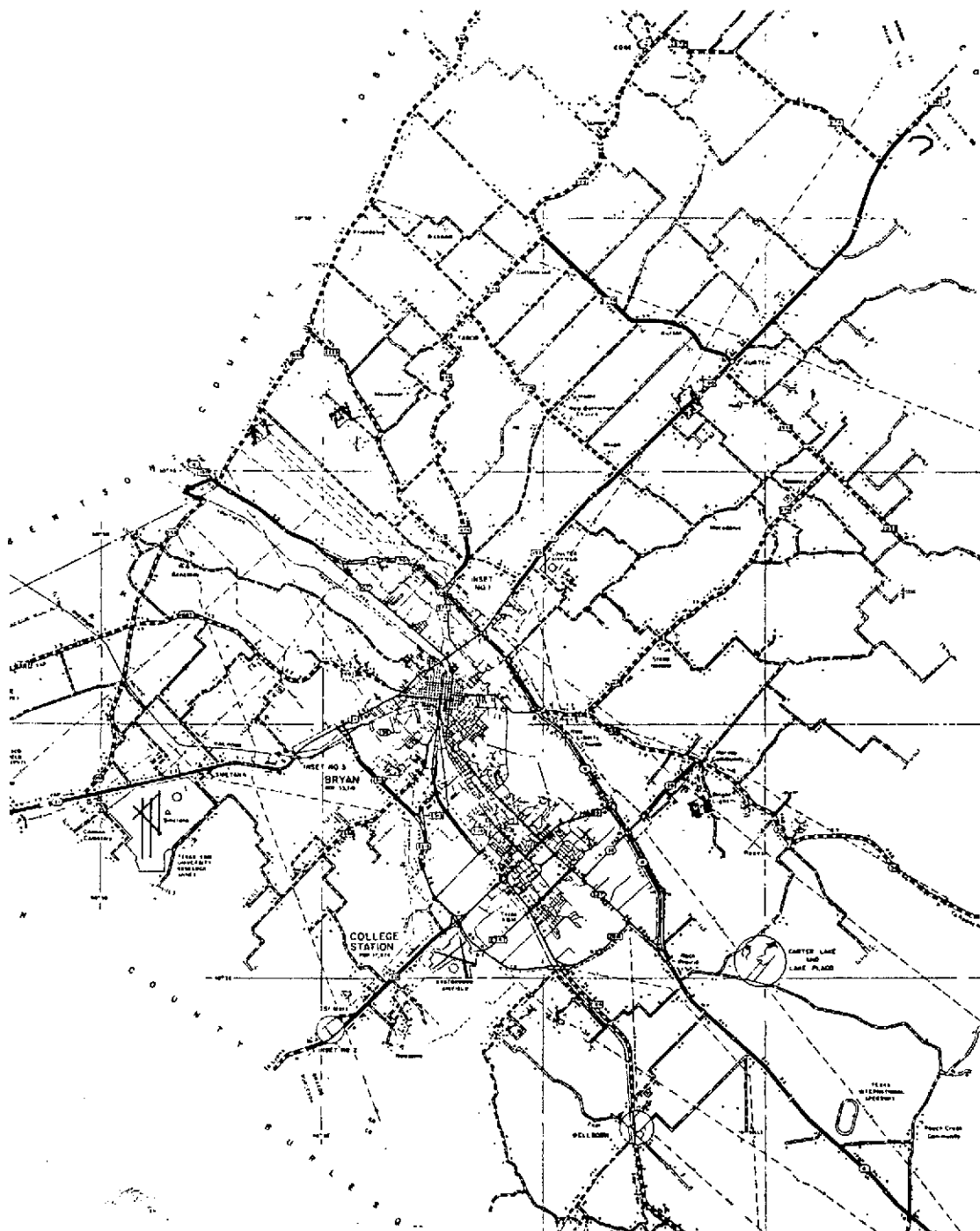


Figure VII-10. General Highway Map of Brazos County

ORIGINAL PAGE IS
OF POOR QUALITY

represented by the attribute values 1-6, respectively. These highways are entered according to coordinates which represent the end point of straight line segments which appropriate the individual highways. A reconstruction of the highway system by IRIMAS from the piecewise linear segments is shown in Figure VII-11. The application of data transformations provides for the selection and display of individual highway systems from the input files. In a like manner, coordinate transformations are applied to modify the scale or the orientation of the IRIMAS reconstruction to align the highway information with resource data from other sources.

This demonstration of INFORM shows the construction of IRIMAS data files and the transformation and reconstruction of data from these files.

Another fruitful source of resource information is the imagery produced by the Earth Resources Technology Satellite. The next section addresses the creation of IRIMAS data files from interactive, IRIMAS-aided interpretation of ERTS imagery from Brazos County.

Image Processing

The software entitled IMAGE is that part of IRIMAS which implements the image analysis functions. This program segment contains the drivers for input and display

\$BRAZOS CO. HIWAY MAP, COORD-96:30W, 30:45N, JAS 1974:
 &A:HWY6=1,HWY6BUS=2,HWY21=3,HWY30=4,HWY60=5,HWY1179=6.

```

0000  A              CC      FF
0001  A              C      F
0002  AAA          C      F
0003  A            C      F
0004  BCRAAA      C      F
0005  BB A        C      F
0006  BB A        C      F
0007  BB A C      F
0008  BB CC      F
0009  C A        F
0010  CB A        F
0011  CCCC B A    F
0012  C          FFFF6FFFF
0013  CC         B A
0014  C          B A
0015  CCC        B A      DDD
0016             B A      D D
0017             B A      D D
0018             B A      D D
0019             B AD     D DDD
0020             EEGB DA   D
0021             E BBD A
0022             E BF A
0023             E BB A
0024             E BB A
0025             E BBA
0026             E C
0027             E A
0028             E A
0029             E A
0030             E A
0031  EEEE        A
0032  EE          AA
0033  E           AA
0034             A
0035             A
0036

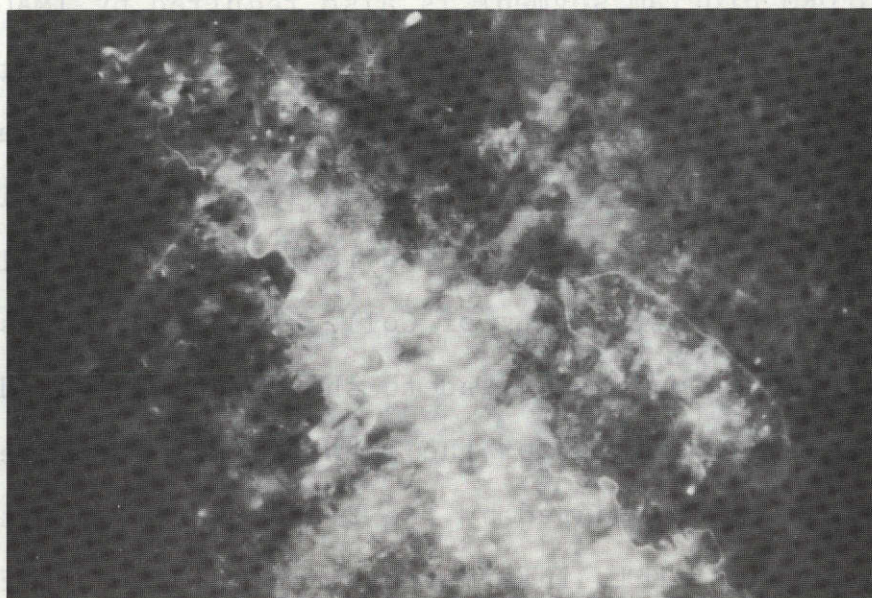
```

Figure VII-11. IRIMAS Reconstruction of Highway System

of ERTS data from digital magnetic tape. The image processing functions described in Chapter V are also implemented. Essentially, the same procedure required for the operator specification of function directives in the INFORM program segment is also required by IMAGE.

The demonstration for the software segment IMAGE utilizes an ERTS scene from Brazos County, Texas. The portion of this scene to be analyzed is shown in Figure VII-12. This scene is of a region approximately 15 miles southwest of the city of College Station, Texas. The demonstration is devised to show typical procedures which might be utilized for the identification of the water resource represented by the Brazos River, and the incorporation of this information into IRIMAS data files. Multiband imagery from ERTS is examined a single band at a time, classification parameters are determined through a training process, and the scene is classified to enhance the resource. The interpreted data is entered into an IRIMAS data file for future reference. In Figure VII-13 is shown the CRT display system from which demonstration results are photographed. Radiance values measured by the ERTS satellite are stored on the digital magnetic tapes supplied by NASA, and these values are displayed in an image format on the CRT system.

of ERTS data from digital magnetic tapes. The image processing functions described in Chapter V are also implemented. Essentially, the same procedure required for the operator specification of function directives in the



Multiband imagery from ERTS is examined a single band at a time. Classification parameters are determined through a training process, and the scene is classified to enhance the resource. The interpreted data is entered into an IRIMAS data file for future reference. In Figure VII-13 is shown the CRT display system from which demonstration results are photographed. Radiance values measured by the ERTS satellite are stored on the digital magnetic tapes

Figure VII-12. ERTS Scene from Brazos County, Texas

**ORIGINAL PAGE IS
OF POOR QUALITY**

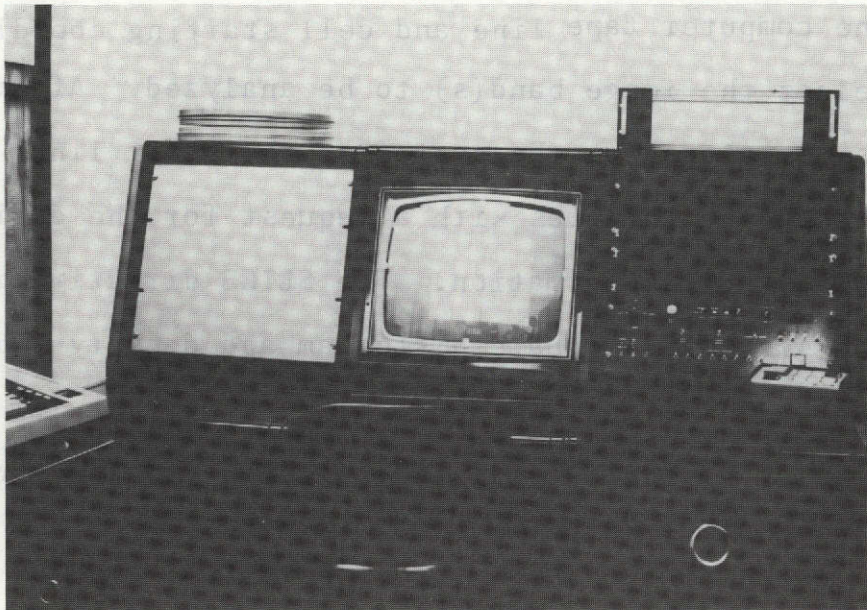


Figure VII-13. CRT Display System

ORIGINAL PAGE IS
OF POOR QUALITY

IMAGE Operator Directives

The operator control sequence for ERTS image data transformation and display is shown in Figure VII-14. Several features of IRIMAS are demonstrated here. The input driver for ERTS requests that the operator supply the computer tape line and cell starting coordinates and select the image band(s) to be analyzed. After the tape is positioned by the system at the start line, the program execution continues with a request for the definition of the data transformation. The SLICE directive requires the system to prepare histogram and density slice information while the HISTOGRAM directive calls for a printer display of the data histogram which is also shown in Figure VII-14.

Image Analysis

Analysis of the ERTS scene typically begins with the preparation of data histograms for each of the ERTS image bands. This analysis step allows the operator to evaluate the range and distribution of the data values. Histograms consequently are prepared for the Brazos Valley scene to be analyzed and are shown in Figure VII-15. The histogram printout gives the total number of points considered, and the number of picture elements at each radiance level. This information is also graphically displayed.

IRIMAS--AN INTERACTIVE RESOURCE
INFORMATION AND ANALYSIS SYSTEM
ENTER ANALYSIS FUNCTIONS

```
0001  ERTS
0002  TRANSFORM
0003  SLICE
0004  DISPLAY
0005  GO TO 1
0006  HISTOGRAM
0007  PAUSE
```

```
ERTS-1 DATA INPUT DEFINE START LINE/CELL
LINES(0-2760),CELLS(0-810)
```

START LINE= 825

START CELL= 0

SELECT BANDS (1,2,3,4,A=ALL)

BAND SELECT= A

TAPE LINE NO. 0825

ENTER TRANSFORM

```
0001 IF (A4.EQ.0) THEN A4=128
```

```
0002  B0=A1+A2+128/A3+128/A4
```

0003 A1=00/16

0004 END

END OF DATA SET ON INPUT UNIT♦♦CONTINUE

HISTOGRAM

TOTAL POINTS

16128

MAXIMUM VALUE

06220

```

000      00000      *
001      00000      *
002      01201      *****
003      06220      *****
004      05856      *****
005      01938      *****
006      00353      **
007      00105      *
008      00056      *
009      00009      *
010      00056      *
011      00176      *
012      00122      *
013      00032      *
014      00004      *
015      00000      *
016      00000      *

```

DEFINED FUNCTIONS ARE COMPLETED♦♦CONT.

Figure VII-14. Operator Control Sequence
for ERTS Processing

ORIGINAL PAGE IS
OF POOR QUALITY

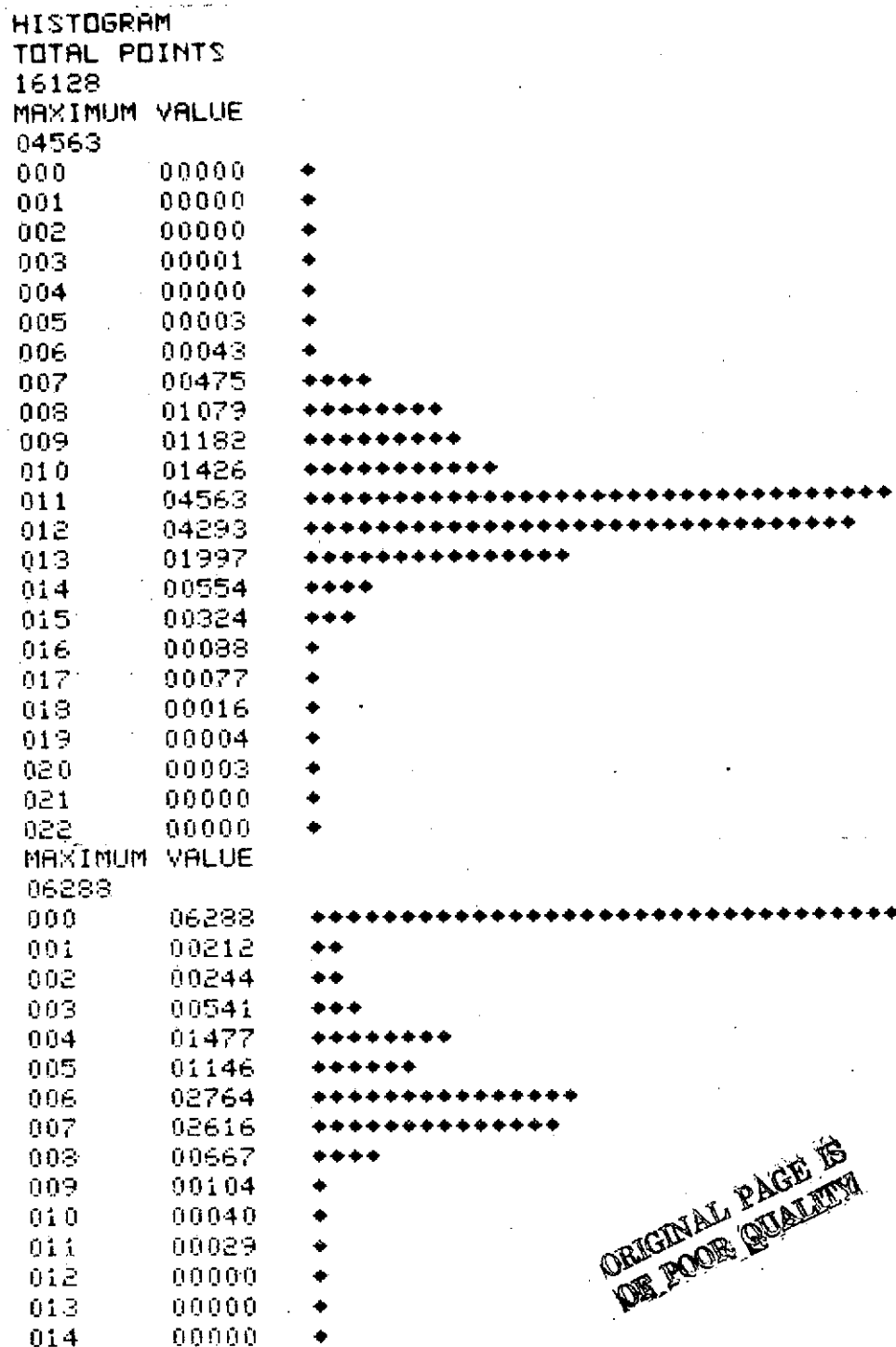


Figure VII-15. Histograms of ERTS Image Bands (cont.)

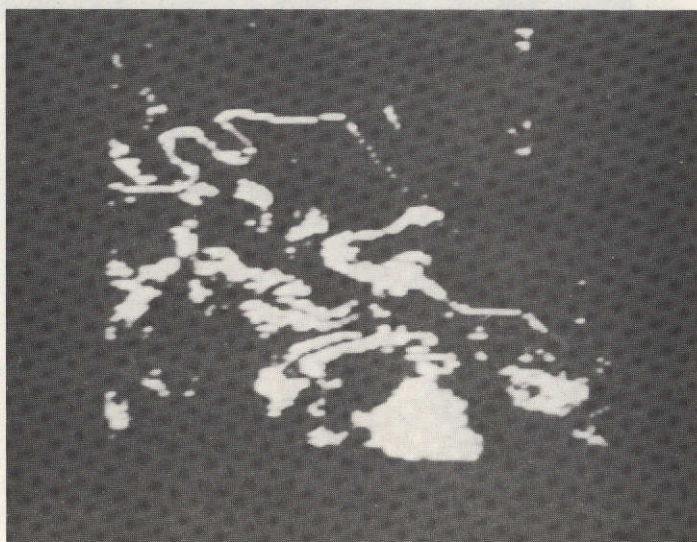
In Figure VII-16 are shown the CRT image displays of bands 1-3, which have been manually enhanced through interaction with the CRT display to separate the radiance levels associated with the river. Band 4 is omitted because of the large amount of 0-level data indicated on the histogram which is attributed to poor data quality.

Maximum likelihood classification provides for the resolution of elements of the scene which correspond to a defined category. In Figure VII-17 is shown the operator sequence required for the training of the classifier software. Upon execution, the system requests the number of classes to be considered, and the coordinate boundaries of the rectangular areas from which the data are to be considered. Names of the data sets are additionally requested. The STATISTICS directive causes the printouts of the results of the training phase. These results for the example are shown in Figure VII-18. The training phase prepares the statistical parameters required by the classification algorithm.

The operator directive sequence for the specification of maximum likelihood classification of ERTS imagery is demonstrated in Figure VII-19. The classifier portion of IRIMAS utilizes the class statistics obtained during the training phase for classification. The result of this classification with only the "water" class displayed is



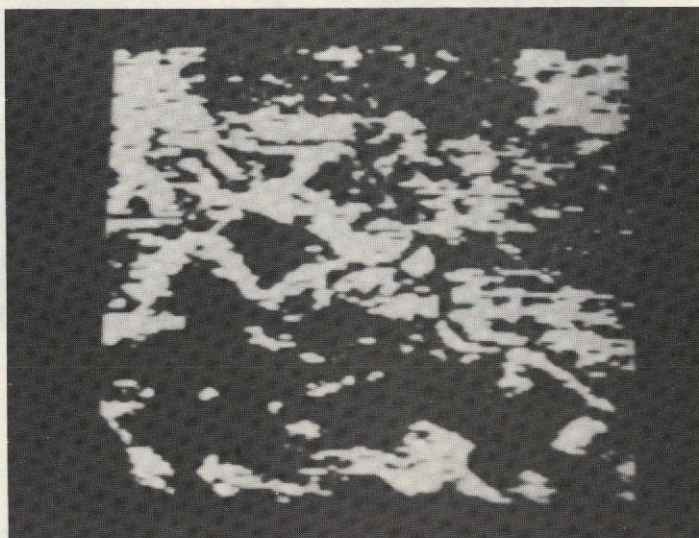
.5-.6 μ Spectral Band



.6-.7 μ Spectral Band

Figure VII-16. Display of ERTS Image Bands

ORIGINAL PAGE IS
OF POOR QUALITY



.7-.8 μ Spectral Band

Figure VII-16. Display of ERTS Image Bands (cont.)

REPRODUCTION PAGE 12

ENTER ANALYSIS FUNCTIONS
 0001 ERTS
 0002 TRAIN
 0003 GO TO 1
 0004 PAUSE
 ERTS-1 DATA INPUT DEFINE START LINE/CELL
 LINES(0-2760), CELLS(0-810)
 START LINE= 825
 START CELL= 0
 SELECT BANDS(1,2,3,4,A=ALL)
 BAND SELECT= A
 TAPE LINE NO. 0825
 CLASSIFICATION TRAINING SETS
 ENTER NUMBER OF CLASSES(1-4) 4

NAME OF CLASS C1:	W1
LEFT-HAND BOUNDARY:	82
RIGHT-HAND BOUNDARY:	92
UPPER BOUNDARY:	94
LOWER BOUNDARY:	96
NAME OF CLASS C2:	W2
LEFT-HAND BOUNDARY:	122
RIGHT-HAND BOUNDARY:	127
UPPER BOUNDARY:	126
LOWER BOUNDARY:	134
NAME OF CLASS C3:	W3
LEFT-HAND BOUNDARY:	138
RIGHT-HAND BOUNDARY:	148
UPPER BOUNDARY:	133
LOWER BOUNDARY:	141
NAME OF CLASS C4:	W4
LEFT-HAND BOUNDARY:	140
RIGHT-HAND BOUNDARY:	157
UPPER BOUNDARY:	174
LOWER BOUNDARY:	188
END OF DATA SET ON INPUT UNIT♦♦CONTINUE	

Figure VII-17. Operator Sequence for Classifier Training

```

CLASS 1
TOTAL POINTS
00033
MEAN VALUES
0032    0029    0048    0025
COVARIANCE
0011
0024    0058
-005    -016    0016
-011    -029    0014    0018
CLASS 2
TOTAL POINTS
00054
MEAN VALUES
0040    0045    0053    0015
COVARIANCE
0019
0036    0075
0022    0033    0058
0038    0066    0081    0198
CLASS 3
TOTAL POINTS
00099
MEAN VALUES
0039    0044    0053    0009
COVARIANCE
0022
0045    0109
0001    -011    0043
-018    -072    0046    0195
CLASS 4
TOTAL POINTS
00270
MEAN VALUES
0039    0047    0050    0013
COVARIANCE
0006
0012    0028
0011    0022    0025
0004    0010    0009    0154

```

Figure VII-18. Classifier Training Results

```

ENTER ANALYSIS FUNCTIONS
0001  ERTS
0002  CLASSIFY
0003  TRANSFORM
0004  DISPLAY
0005  GO TO 1
0006  PAUSE
ERTS-1 DATA INPUT DEFINE START LINE/CELL
LINES(0-2760), CELLS(0-810)
START LINE= 825
START CELL= 0
SELECT BANDS(1,2,3,4,A=ALL)
BAND SELECT= 2
TAPE LINE NO. 0825
ENTER TRANSFORM
0001  IF(A1.NE..
                2) THEN A1=0
0002  IF(A1.EQ.2) THEN A1=15
0003  END
END OF DATA SET ON INPUT UNIT♦♦CONTINUE
DEFINED FUNCTIONS ARE COMPLETED♦♦CONT.
ENTER ANALYSIS FUNCTIONS
0001

```

ORIGINAL PAGE 3
FOR FOOT QUALITY

Figure VII-19. Operator Specification of Classify Sequence

shown in Figure VII-20. At this point in the analysis, the water resource identified as the Brazos River through interpretation of the image presentation is interactively saved in the IRIMAS data files. In Figure VII-21 is shown the listing of this data file.

Change Monitoring

A third demonstration is used to illustrate the versatility of the IRIMAS data system in the monitoring of resource changes over a period of time. Monitoring of changes using the traditional grid format storage would be exceedingly difficult especially with the overlay and registration of multi-temporal images. Use of the IRIMAS data formats and analysis capabilities make this monitoring a simple interactive task.

The feature chosen to show a procedure for change analysis is the west-bypass of U.S. Highway 6 around the Bryan-College Station, Texas communities. Earth Resources Technology Satellite data from the area has been accumulated for the period August 1972 through April 1973, a period during which the bypass has shown considerable development. The latest available general highway map of the area (see Figure VII-10, page 190) does not show the west bypass, however, construction appears on the recent ERTS imagery. In Figure VII-22, are shown several image

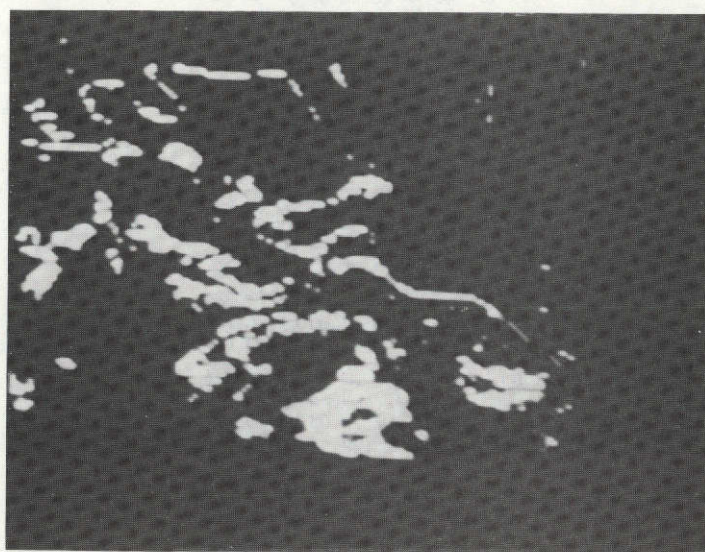


Figure VII-20. Classification Results for "Water" Class

ORIGINAL PAGE IS
OF POOR QUALITY

ORIGINAL PAGE IS
OF POOR QUALITY

\$BRAZOS WATR,BIAS TPLINE 825,CELL 0,DATA BIAS 64,64,ERTS

&A:15//

♦LBR1#A15%X71Y99A%X77Y108A%X82Y112A%X87Y112A%X90Y108A%X90Y106A%X84Y99A
♦LBR2#A15%X84Y99A%X84Y97A%X91Y96A%X97Y102A%X102Y103A%X105Y101A%X102Y96A
♦LBR3#A15%X102Y96A%X97Y92A%X99Y90A%X104Y90A%X111Y91A%X120Y92A%X128Y91A
♦LBR4#A15%X128Y91A%X126Y91A%X136Y104A%X142Y114A%X144Y119A%X140Y123A%X132Y125A
♦LBR5#A15%X132Y125A%X124Y127A%X132Y130A%X127Y136A%X133Y137A%X141Y132A%X146Y131A
♦LBR6#A15%X146Y131A%X150Y134A%X150Y140A%X156Y146A%X171Y147A%X186Y157A%X197Y166A
♦AAR1#A15%X134Y176A%X140Y177A%X148Y170A%X162Y189A%X141Y189A%X134Y178A%X134Y176A
♦AAR2#A15%X135Y139A%X140Y134A%X147Y134A%X147Y138A%X150Y138A%X150Y141A%X138Y141A
/♦

Figure VII-21. IRIMAS File Listing of Brazos River Data

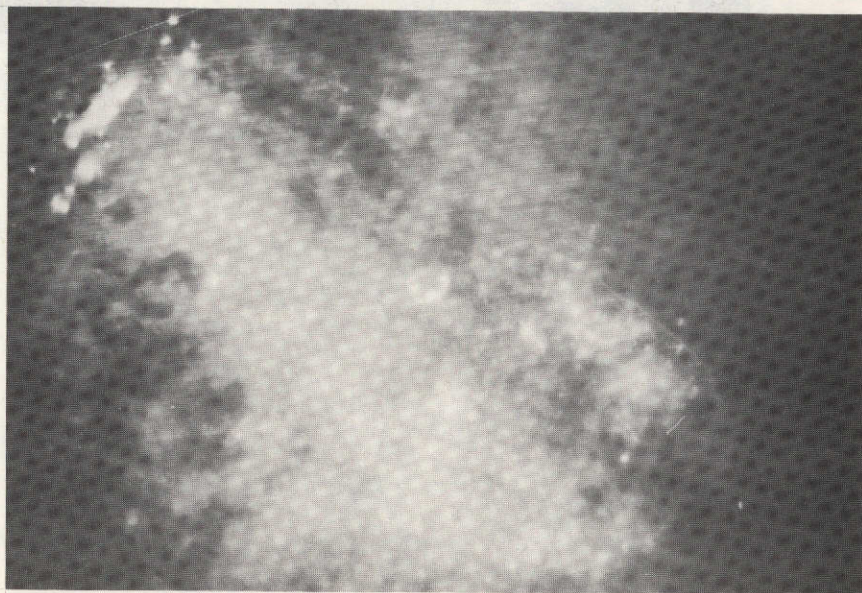
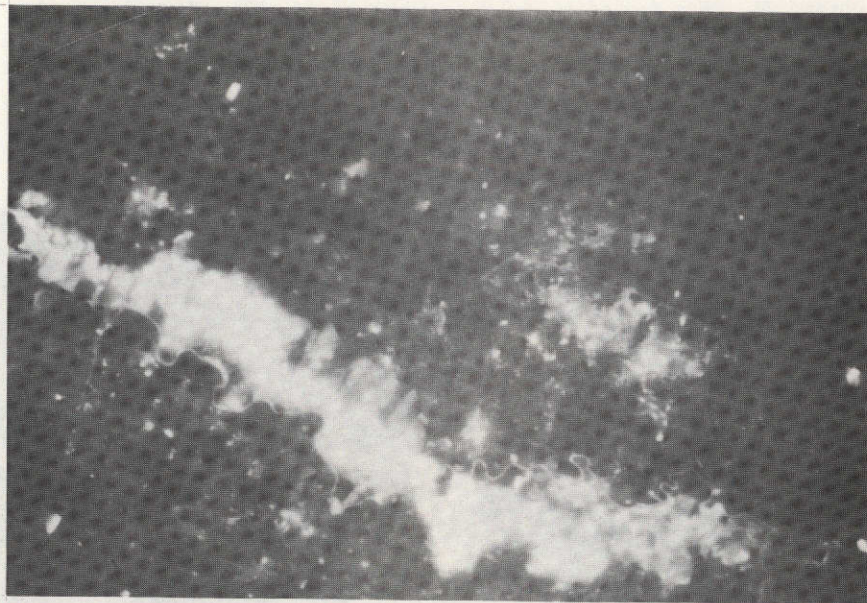


Figure VII-22. ERTS Image Sequence During Highway 6 West Bypass Construction

ORIGINAL PAGE IS
OF POOR QUALITY

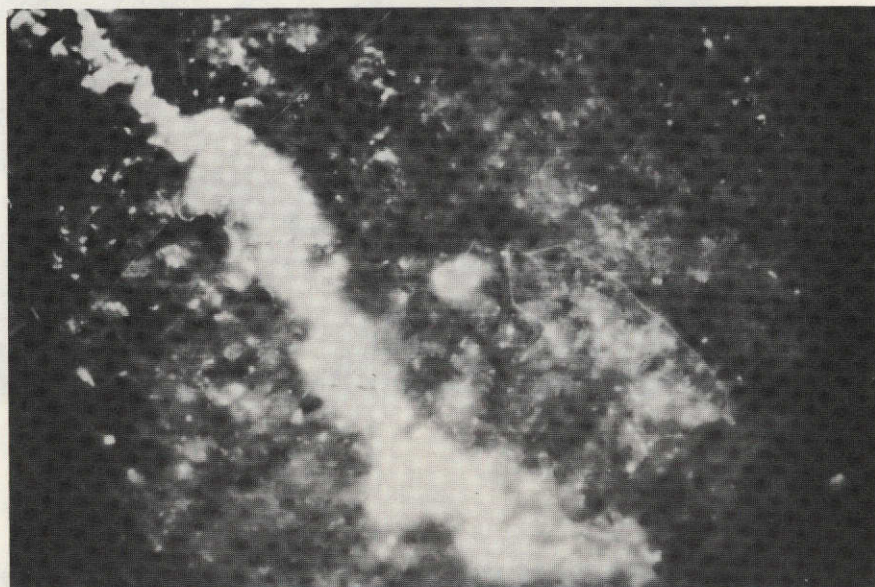


Figure VII-22. ERTS Image Sequence During Highway 6 West Bypass Construction (cont.)

ORIGINAL PAGE IS
OF POOR QUALITY

reproductions of the area under consideration for the period where development of the bypass is apparent. Bypass changes are interpreted from the imagery and are added to the IRIMAS data file created for the general Brazos County highway system. An IRIMAS expansion of the data file is displayed in Figure VII-23.

The sequence of highway construction is indicated by different letters on the printout. The "*" symbol represents the major highway systems which occur on the general highway map. Bypass changes interpreted from ERTS imagery are encoded by date as follows: 1) B-5/27/73, 2) C-7/20/73, 3) G-7/20/73, 4) D-3/29/74.

Maintenance of the interpreted information requires minimal effort in IRIMAS data files where changes in resource information can be readily saved. Reconstruction of these data is also provided for by the IRIMAS system. Traditional data formats and image interpretation systems preclude the simple maintenance of the interpreted data in a single file. Full images must be maintained and overlaid for each date considered. The overlay method also requires full interpretation of the image for each analysis and does not take advantage of previous interpretations, whereas IRIMAS provides for accurate recall of past interpretations.

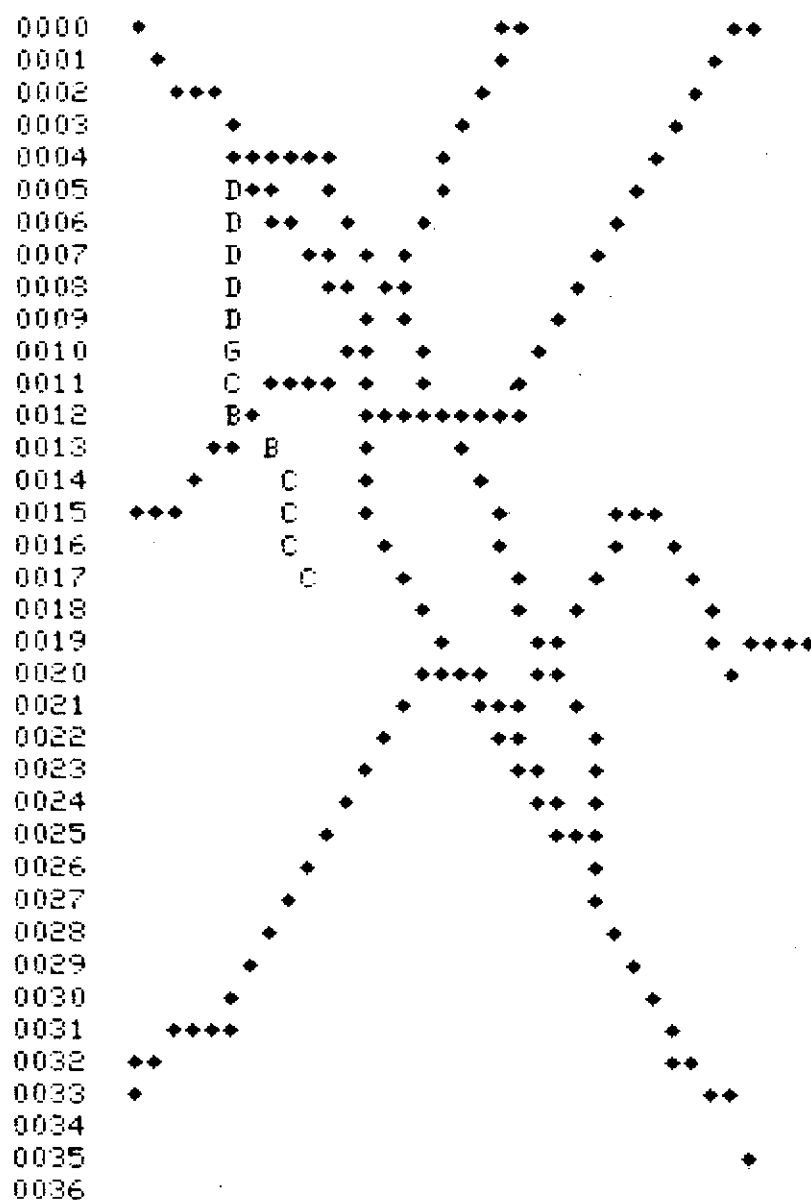


Figure VII-23. New Highway Construction Update
to General Highway Map

CHAPTER VII

CONCLUSIONS AND RECOMMENDATIONS

The recent availability of timely synoptic earth imagery from the Earth Resources Technology Satellite (ERTS) provides a wealth of information for the monitoring and management of vital natural resources. Satellite data, as well as information derived from aircraft sensors, are providing an increasingly more important supplement to conventional data sources used for the monitoring, protection, and allocation of resources. Remote sensor data provides a unique point of view which enhances the information obtained about an area resource from traditional ground survey sources. It is becoming apparent, due to the scope and amount of data available, that computer processing and analysis of resource information will become a necessity for timely application. Computer software systems have not existed to provide the planning function with an adequate information structure to maintain and utilize spatial and temporal data interpretations. Available systems, for the most part, also lack the simplistic interface required by the resource planner who is not versed in the complexities of computer processing.

The resource management decision is concerned with the evaluation of extrapolated land usage and its

associated environmental impact, both in an assumed steady state and in the presence of actual or anticipated perturbations. The resource manager asks composite questions about the relationships of area extensive resource features. These questions amount to a conceptual overlay of any of the "maps" which comprise the data base. Decisions are based on the disposition of related resources such as soils, water, and vegetation.

The objective of this research has been to develop an information management system suited to the general needs of resource managers engaged in the utilization of imagery and other spatiotemporal data in their management processes. The need for such a system has been inferred by numerous authors [1], [2], [6], [26] who have emphasized the need for an interactive analysis system capable of handling the information requirements of the spatial and temporal data available from a variety of sources.

Several processing systems have been developed to address one aspect or another of the resource management problem, however each of these fails with respect to one or more of the essential ingredients for an effective management tool. These ingredients include a simplistic, interactive interface; the flexibility to handle a variety of information types; and a system whose costs are in-line with operating budgets at a regional agency level. The

Interactive Resource Information Management and Analysis System (IRIMAS) has been developed to specifically address the need for each of these ingredients in the handling and analysis of the spatial and temporal information required for resource related decision-making. The system provides for the interactive interpretation of data from ERTS imagery, cartographic products, and point surveys. Operator interface is simplistically accomplished through single word commands and computer initiated queries and requires only a minor amount of operator training for its effective use. A flexible data base has been set up to address point, line and area format data which can be retrieved for the display of overlays according to user priority-weights. Additionally, the system can be duplicated for equipment costs below \$50,000, or may be set up in an interactive time-shared mode to take full advantage of existing machines.

IRIMAS is innovative from several significant aspects. The application of formal language theory to the management of information data files has been alluded to only recently in the literature, yet it provides the basis for the flexible information format developed for this system. The transformation and maintenance of interpreted sensor information in data files, compressed by the utilization of category boundaries, offers a worthwhile alternative to

the accepted processing standards prevalent in image analysis today. The greatest significance of IRIMAS, however, lies in the evolution of a new approach to the utilization of remote sensor data and computer processing techniques in resource analyses and in the introduction of a new tool which exhibits a great amount of flexibility and potential for interactive interpretation of resource information.

The Interactive Resource Information Management and Analysis System, IRIMAS, is a hardware/software system developed with a Texas Instruments 980A minicomputer system consisting of a CPU with 12K words of memory, a 9-track digital magnetic tape unit, a send-receive terminal with dual magnetic tape digital cassettes, and a color CRT image display. Several additional pieces of equipment would suitably augment this hardware arrangement. Experience shows that the 12K words of memory is inadequate to handle the fully implemented IRIMAS software package. It is recommended that future configurations of IRIMAS can be provided with an additional 8K words of memory to add system flexibility, and to remove some of the software restrictions placed on the system.

The information storage format requires the definition of resource features in terms of coordinate boundaries. It is time-consuming and cumbersome to manually

extract the coordinate information and to enter it into the data files. This defining operation could be more efficiently accomplished with the aid of an automated graphics input device. Several varieties of these instruments are available commercially and would greatly enhance the efficiency of IRIMAS utilization. Future implementation of IRIMAS could effectively utilize this class of input device, and its inclusion in the system is recommended.

Several inadequacies are noted in the IRIMAS software and modifications are recommended. The cluster algorithm implemented in IRIMAS has produced less than desirable results in the analysis of ERTS imagery. Several reasons are suggested for this lack of performance. Primarily, the difficulty seems to lie in the number of classes permitted, which has been limited to sixteen, because of memory constraints. Imagery from ERTS characteristically has more individual classes associated with a scan line than was experienced with data sets in previous applications of the algorithm. These data differences suggest that cluster thresholds should also be examined for the ERTS data.

The maximum likelihood classification algorithm applied in IRIMAS, has a fixed threshold for the identification of the "no classification" category which has been

set high enough to allow classification of all points. It is recommended that this threshold be set by the system operator with the specification of percentage of points to be included or rejected by the algorithm. Computation or table look-up could be applied for the determination of the appropriate threshold as a function of the number of dimensions and individual class statistical parameters. This procedure would facilitate the single category classification of a scene with the rejection of other categories which would be useful in a resource interpretation system.

Coordinate and data transformation software is implemented to support fixed-point binary arithmetic. This method was implemented to take advantage of the hardware integer multiply/divide feature of the TI980A system. Operational use of the system has indicated that the transformations normally required would be more flexible and precise if arithmetic operations were performed by floating-point operations rather than the fixed-point binary operations. It is recommended therefore that the transformation interpreter be modified to perform floating-point arithmetic rather than fixed-point arithmetic.

Operational use of the IRIMAS classification algorithms is extremely slow computationally. Classification of the 128 x 128 pixel scene using a single ERTS band requires in excess of 3 minutes, and use of the full four

bands expands the time to 12-15 minutes. These times are undesirable from an interactive point of view and could be significantly reduced if the classification software were converted to assembly language rather than FORTRAN.

Other system improvements could be accomplished by the incorporation of software drivers to allow inclusion of additional image format data sources, and to include image spatial-frequency processing. With the addition of a disc to the TI980A system, drivers could be developed to use the sector addressable capability of that system to enlarge the processing window by writing to the disc rather than into computer memory. Overlay link-editing and real-time loading procedures could be applied in the disc system to improve the efficiency of memory usage. Additionally, it seems practical to develop the software to allow utilization of the CRT graphics-overlay feature to superimpose resource information graphically into an image display. These recommended systems improvements should add considerable versatility to the system and should enable more effective use of the system as an interactive resource management aid.

REFERENCES

- [1] Dueker, K. J. and F. E. Horton, "Utility of Remote Sensing Data for Urban Land Use Planning", Proc. of Seminar in Remote Sensing, American Society of Photogrammetry, Houston, Texas, February 1972, pp. 178-190.
- [2] Brooner, W. G. and D. A. Nichols, "Considerations and Techniques for Incorporating Remotely Sensed Imagery Into the Land Resources Management Process", Remote Sensing of Earth Resources, Vol. 1, F. Shahrokhi ed., Univ. of Tenn., Tullahoma, Tenn., 1972, pp. 1-24.
- [3] Swanlund, G., "Hybrid Techniques for Automatic Interpretation", Proc. of Seminar on Remote Sensing, American Society of Photogrammetry, Houston, Texas, February 1972, pp. 245-256.
- [4] Viglione, S. S., "Digital Data Processing and Analysis", Proc. of Seminar on Remote Sensing, American Society of Photogrammetry, Houston, Texas, February 1972, pp. 208-244.
- [5] Andrews, H. C., Computer Techniques in Picture Processing, New York: Academic Press, 1970.
- [6] Nagy, G., "Digital Image-Processing Activities in Remote Sensing for Earth Resources", Proc. IEEE, Vol. 60, No. 10, October 1972, pp. 1177-1200.
- [7] Kanal, L. N., "Interactive Pattern Analysis and Classification Systems: A Survey and Commentary", Proc. IEEE, Vol. 60, No. 10, October 1972, pp. 1200-1215.
- [8] Rees, H.J.B., "Time Series Analysis and Regional Forecasting", Regional Forecasting, M. Chisholm, A. Frey, P. Haggett, ed., Proc. of the 22nd Symp. of the Colston Res. Soc., University of Bristol, April 1970, pp. 11-24.
- [9] Cordey-Hayes M., T. A. Broadbent and D. B. Massey, "Towards Operational Urban Development Models", Proc. of the 22nd Symp. of the Colston Res. Soc., University of Bristol, April 1973, pp. 179-220.
- [10] Hagerstrand, T., "Regional Forecasting and Social Engineering", Proc. of the 22nd Symp. of the Colston

- Res. Soc., University of Bristol, April 1974, pp.1-10.
- [11] Harris, B., "New Tools for Research and Analysis in Urban Planning", Urban Planning in Transition, E. Erber, ed., Grossman, New York, 1970, pp. 192-202.
 - [12] Candeub, T., "New Techniques in Making the General Plan", Urban Planning in Transition, E. Erber, ed., Grossman, New York, 1970, pp. 216-224.
 - [13] Satsangi and Ellis, "A Generalized System Theoretic Framework for Modeling Large Scale National Economics Systems in Dynamic, Structural, and Spatial Terms", Int. Jour. of Sysm. Sci., Vol. 2, Sept. 1971.
 - [14] Braswell, "A Systems Approach to the Microresource Reallocation and Control Problem", Int. J. Sysm. Sci., Vol. 2, No. 3, 1971, pp. 225-230.
 - [15] Brooner, W. and D. A. Nichols, "Considerations and Techniques for Incorporating Remotely Sensed Imagery into the Land Resources Management Process," Remote Sensing of Earth Resources, Vol. 1, F. Shahrokhi, ed., Univ. of Tenn., Tullahoma, Tennessee, 1972, pp. 1-24.
 - [16] Thomlinson, R., Urban Structure - The Social and Spatial Character of Cities, Random House, Inc., 1969, pp. 28-29.
 - [17] Rouse, J. W., Jr. and J. A. Schell, "On a Systems Approach to Earth Observations", National Telecommunications Conference, Houston, Texas, December 1972.
 - [18] Zaitzeff, E. M., C. L. Kach and C. L. Wilson, "MSDS: An Experimental 24-Channel Multispectral Scanner System", IEEE Trans. GeoSci, Elec., Vol. GE-9, No. 3, July 1971, pp. 114-120.
 - [19] O'Handley, D. and W. B. Green, "Recent Developments in Digital Image Processing at the Image Processing Laboratory at the Jet Propulsion Laboratory", Proc. IEEE, Vol. 60, July 1972, pp. 821-828.
 - [20] Pratt, E. K., "Semi-Annual Technical Report of the USC Image Processing Laboratory", Univ. of South California, Rep. 411, March 1972.
 - [21] Landgrebe, D. A., P. J. Min, P. H. Swain, and K. S. Fu, "The Application of Pattern Recognition Techniques

to a Remote Sensing Problem", Purdue University, Lafayette, Indiana, 1968, LARS Information Note 080568.

- [22] Marshall, R. E. and F. J. Kreigler, "Multispectral Recognition Systems", Proc. 9th IEEE Symp. Adaptive Processes, Decision, and Control, Austin, Texas, p. XL-2, 1970.
- [23] Haralik, R. M. and D. E. Anderson, "Texture-tone Study with Application to Digitized Imagery", Univ. of Kansas, CRES Tech. Rep. 182-2, 1971.
- [24] Eppler, W. G., D. L. Loe, E. L. Wilson, S. L. Whitley, and R. J. Sachen, "Interactive System for Remote Sensor Data Analysis", Proc. 7th Int. Symp. Remote Sensing Environ., Ann Arbor, Michigan, 1971, pp. 1293-1306.
- [25] Borden, F. Y., "A Digital Processing and Analysis System for Multispectral Scanner and Similar Data", Remote Sensing of Earth Resources, Vol. 1, F. Shahrokhi, ed., The University of Tennessee, Tullahoma, Tenn., 1972, pp. 481-507.
- [26] Phillips, R. L., "Computer Graphics in Urban and Environmental Systems", Proc. IEEE, Vol. 62, No. 4, April 1972, pp. 437-452.
- [27] Advanced Scanners and Imaging Systems for Earth Observations, NASA SP-335, December 1972, pp. 39.
- [28] Swain, P. H., "Data Processing I: Advancements in Machine Analysis of Multispectral Data", Purdue Univ., LARS Information Note 012472, 1972.
- [29] Martin, J. T., Design of Man-Computer Dialogues, Prentice Hall, 1973, Englewood Cliffs, New Jersey, p. 3.
- [30] Martin, J. T., Design of Man-Computer Dialogues, Prentice Hall, 1973, Englewood Cliffs, New Jersey, pp. 449-450.
- [31] Hopcroft, J. E. and J. D. Ullman, Formal Languages and Their Relation to Automata, Addison-Wesley Publishing Company, Inc., Reading, Mass., 1969, pp. 1-25.

- [32] Book, R. V., "Topics in Formal Language Theory," Currents in the Theory of Computing, A. V. Aho, ed., Prentice Hall, Englewood Cliffs, New Jersey, 1973, pp. 1-34.
- [33] Gauthier, R. L. and S. D. Ponto, Designing Systems Programs, Prentice Hall, Inc., Englewood Cliffs, New Jersey, 1973, pp. 83-97.
- [34] Gauthier, R. L. and S. D. Ponto, Designing Systems Programs, Prentice Hall, Inc., Englewood Cliffs, New Jersey, 1974, pp. 68-70.
- [35] Newman, W. M. and R. F. Sproul, Principles of Interactive Computer Graphics, McGraw-Hill Book Company, pp. 123-125.
- [36] Andrews, H. C., Computer Techniques in Image Processing, Academic Press, New York, New York, 1970, pp. 73-112.
- [37] Schell, J. A., "A Comparison of Two Approaches for Category Identification and Classification Analysis from an Agricultural Scene", Remote Sensing of Earth Resources, Vol. 1, F. Shahrokhi, ed., The Univ. of Tenn., Tullahoma, Tenn., 1972, pp. 374-394.
- [38] Bruns, F. J., V. T. Rhyne and J. A. Schell, "A Color CRT Image Display System", Proc. of Int. Symp. of Society on Information Display, New York, May 1973.
- [39] Schell, J. A., "Software Listings for IRIMAS", Technical Memorandum RSC-95, Remote Sensing Center, Texas A&M University, College Station, Texas, August 1974.
- [40] Woodlands New Community - Guidelines for Site Planning, McHargm Roberts and Todd, Philadelphia, Penn., December 1973.
- [41] Ecological Planning Study for the New Community, Wallace, McHarg, Roberts and Todd, Houston, Texas, July 1971.
- [42] Haas, R.H. and M.C. McCaskill, "Use of Large Scale Aerial Photography in Obtaining Vegetation Information for Urban Planning," Technical Report RSC-39, Texas A&M University, Remote Sensing Center, August 1972.